

omzien naar weldoen

Tabel.

8	9						
Of het gesticht geest. redicant wordt en toe ja door men en met haer vel?	Beveling, doch alleen van de kerkzinnige personen. Niet uitsluiting van diegene welke om wonderdraag opgevoerd kunnen staan. Het is een juist, als mede niet uitsluiting van alle kerkzinnige personen die in het gesticht mochten wonen, blyven ten, verdunde 8. f. verdunde in 	N. SIJZEN			Solaal	Rijk van een kerkzinnige kerkzinnige of door een kerkzinnige werd kerkzinnige kerkzinnige geest. geest ten kerkzinnige kerkzinnige en den kerkzinnige kerkzinnige onder kerkzinnige kerkzinnige	

Omzien naar Weldoen

*... ook een uurwerk immers dat slechts
uit veren en tandwielen bestaat,
kan veel beter dan wij,
met al ons inzicht,
de uren tellen en de tijd meten.*

DESCARTES, Over de methode (1637)

OMZIEN NAAR WELDOEN

Programma-evaluatie in theorie en praktijk

PROEFSCHRIFT

TER VERKRIJGING VAN DE GRAAD VAN DOCTOR
IN DE GENEESKUNDE AAN DE
KATHOLIEKE UNIVERSITEIT TE NIJMEGEN
OP GEZAG VAN DE RECTOR MAGNIFICUS
PROF. DR. B. M. F. VAN IERSEL
VOLGENS BESLUIT VAN HET COLLEGE VAN DECANEN
IN HET OPENBAAR TE VERDEDIGEN OP
DONDERDAG 22 JANUARI 1987,
DES NAMIDDAGS TE 3.30 UUR

DOOR

Arnold Elisabeth Sevaristus Sijben

geboren te Hoensbroek

druk: B.V. Boekdrukkerij F.E. MacDonald – Nijmegen
Instituut Psychiatrie K U. – Publikatie Nr. 1

CIP-GEGEVENEN KONINKLIJKE BIBLIOTHEEK, DEN HAAG

Syben, Arnold Elisabeth Sevaristus

Omzien naar weldoen programma-evaluatie in theorie en
praktijk / Arnold Elisabeth Sevaristus Syben – [S l
s n]

Proefschrift Nijmegen – Met lit opg

ISBN 90-9001479-9

SISO 607 2 UDC 615 851 001 12(043 3)

Trefw evaluatietechniek , psychotherapie

© (1986) A E S Syben

Niets uit deze uitgave mag worden verveelvoudigd en/of openbaar gemaakt door middel van druk,
fotokopie, microfilm of op welke andere wijze ook, zonder voorafgaande schriftelijke toestemming van
de auteur

Inhoudsopgave

Inleiding: de Vraagstellingen	1
Deel A: Programma-evaluatie in Theorie	7
Hoofdstuk 1: Omzien naar weldoen	9
1.1 Inleiding	9
1.2 De psychiatrische “zorg” vóór 1800	10
1.3 Het begin van staatsbemoeienis: 1800-1840	11
1.4 De wet van 1841	18
1.5 Het staatstoezicht van 1842 tot 1884	21
1.6 De periode 1885-1936	27
1.7 Van 1969 tot en met 1984	32
1.8 Van verleden naar toekomst	37
Hoofdstuk 2: Is evalueren ook onderzoeken?	39
2.1 Inleiding	39
2.2 Wat is onderzoek?	41
2.3 Het verschil tussen onderzoek en evaluatie	45
2.4 Vormen van behandelingsevaluatie	51
2.4.1 Summatieve of formatieve evaluatie	51
2.4.2 “Process” of “outcome” evaluatie	53
2.4.3 Programma-evaluatie	54
Hoofdstuk 3: Programma-evaluatie, wat is dat?	56
3.1 Inleiding	56
3.2 Historie en definitie	56
3.3 Vormen van programma-evaluatie	60
3.3.1 De evaluatievragen	61
3.3.2 Organisatievormen van programma-evaluatie	65
3.4 Hoofdlijnen van een programma-evaluatieproject	68
3.4.1 Het beleidsplan	70
3.4.2 Het evaluatieplan	71
3.4.3 De implementatie	71
3.4.4 De evaluatie en feedback	72
3.4.5 Een nieuw beleidsplan	73
3.5 Evalueren en geëvalueerd worden	73
Hoofdstuk 4: Het meten van therapie-effecten	77
4.1 Inleiding	77
4.2 Het operationalisatieprobleem	78

4.3	Geïndividualiseerde doelevaluaties	81
4.3.1	Target Complaints	82
4.3.2	Goal Attainment Scaling	83
4.3.3	Andere doel-evaluatie methoden	85
4.4	Psychologische tests	86
4.5	Subjectieve beoordelingen	87
4.6	Conclusie	90

Deel B: Programma-evaluatie in praktijk	91
--	-----------

Hoofdstuk 5: Kliniek Overwaal	93
5.1 Inleiding	93
5.2 Doelstelling	93
5.3 Voorgeschiedenis	94
5.4 Werkwijze	95
5.5 De behandelingsfasen	96
5.5.1 Het intake-proces	96
5.5.2 Van opname tot behandelingsplan	97
5.5.3 De behandeling	98
5.5.4 Het ontslag	99
5.6 Overwaal als object van programma-evaluatie	100

Hoofdstuk 6: Project Overwaal programma-evaluatie (P.O.P.E.)	101
6.1 Inleiding	101
6.2 Voorgeschiedenis	101
6.3 Doelstelling	103
6.4 De opzet	103
6.5 De evaluatie-fasen	106
6.5.1 De periode 1974 tot 1977	106
6.5.2 De periode 1977 tot 1981	107
6.5.3 De periode 1980 tot 1983	109
6.5.4 De periode 1983-1985	117
6.6 Het einde van P.O.P.E.	120

Hoofdstuk 7: De kwantitatieve gegevens geregistreerd	123
7.1 Inleiding	123
7.2 Algemene registratie gegevens 1974 t/m 1984	123
7.2.1 Opnameduur	125
7.2.2 Demografische variabelen	128
7.3 Problematiek vóór opname	133
7.4 Hulpverlening vóór opname	138

Hoofdstuk 8: De evaluatie van de resultaten	141
8.1 Inleiding	141
8.2 Follow-up onderzoek 1974-1976	142
8.3 De periode november 1976-1980	145
8.3.1 Ontslagvragenlijst patiënt	147
8.3.2 Ontslagvragenlijst therapeut	149
8.3.3 Follow-up patiënt	151
8.3.4 Follow-up nazorg	154
8.4 Conclusie	155
8.5 De periode 1981-1985	156
8.5.1 Opnamevragenlijst patiënt	157
8.5.2 Ontslagvragenlijst patiënt	158
8.5.3 Ontslagvragenlijst therapeut	162
8.5.4 Follow-up vragenlijsten patiënt	165
8.5.5 Follow-up vragenlijsten hulpverlener	168
8.6 De periode 1976-1985	171
8.7 Evaluatie van doelstellingen	174
 Hoofdstuk 9: De resultaten geanalyseerd	 177
9.1 Inleiding	177
9.2 Betrouwbaarheid	178
9.2.1 Interne consistentie	178
9.2.2 Interbeoordelaarsbetrouwbaarheid	180
9.3 Reductie in factoren	181
9.3.1 Ontslagvragenlijst patiënt	181
9.3.2 Ontslagvragenlijst therapeut	182
9.3.3 Follow-up patiënt	183
9.3.4 Follow-up hulpverleners	184
9.4 Samenvatting	184
9.5 Analyse van effecten	186
9.5.1 Psychotherapeut	187
9.5.2 Diagnose	189
9.5.3 Ontslagjaar	191
9.5.4 Andere effecten	191
9.6 Niet-effectieve patiënten	192
9.7 Conclusies	194
 Hoofdstuk 10: Met verschillende maten gemeten	 195
10.1 Inleiding	195
10.2 Inhoudelijke vergelijking	196
10.3 I.D.E. – een idee fixe?	198
10.4 De maten vergeleken	200
10.4.1 De normatieve evaluatie met behulp van standaardlijsten	200
10.4.2 Inhoudvalidering	201
10.4.3 Criterium georiënteerde evaluatie	205

10.4.4	Intercorrelaties effectmetingen	207
10.5	Conclusie	208
Hoofdstuk 11: Programma evaluatie geëvalueerd		211
11.1	Inleiding	211
11.2	De doelstellingen	211
11.3	Programma-evaluatie en organisatieverandering	213
11.4	Overige evaluatie thema's	219
11.5	Aanbevelingen	221
11.6	Tenslotte	224
Samenvatting		225
Summary		241
Bibliografie		257

Voorwoord

In dit proefschrift wordt bepleit dat moet worden “omgezien” naar het “weldoen” in de geestelijke gezondheidszorg. Het is een verslag van een programma-evaluatie-project dat gedurende ruim 8 jaar is uitgevoerd in kliniek Overwaal te Lent. Wanneer ik nu, bij de afronding van dit project, zelf omzie naar deze periode, dan realiseer ik me eens te meer dat dit onderzoek niet uitgevoerd had kunnen worden en deze dissertatie niet tot stand had kunnen komen zonder de bijdrage van zeer velen. Onder begeleiding van mijn twee promotoren, Prof.dr. A.A. Fischer en Prof.dr. J.H. Dijkhuis is dit proefschrift geworden tot wat nu voor u ligt. Prof.dr. S.J. Nijdam was bij mijn indiensttreding hoofd van de afdeling psychiatrie. Hij maakte mogelijk, dat ik op verzoek van Dr. J.W.G. Orlemans, chef de clinique van Overwaal, aldaar werd gedetacheerd. Bij de benoeming van Prof.dr. A.A. Fischer in 1977, werd ik onder zijn leiding geplaatst. In overleg met hem is in 1979 een onderzoeksplan opgesteld, dat tot deze dissertatie heeft geleid. Het Praeventiefonds heeft gedurende 5 jaar middels een subsidie dit onderzoek financieel mogelijk gemaakt.

In dit onderzoek zijn de gegevens van ca. 800 patiënten verwerkt. Hun medewerking was onontbeerlijk. Ieder van hen heb ik steeds persoonlijk ingelicht over dit onderzoek en bedankt voor hun medewerking. Op deze plaats rest mij een dankzegging aan hen allen, en ik wens hen alle goeds toe.

De medewerkers van Overwaal: chef de clinique, psychotherapeuten, sociotherapeuten, creatief en bewegingstherapeuten – te velen om allen met naam te noemen – hebben elk op hun eigen wijze een bijdrage geleverd bij het verrichten van dit onderzoek. Ook de administratieve en huishoudelijke medewerkers wil ik expliciet noemen. Zij verzorgden mede de (im-)materiële randvoorwaarden. Ik heb van 1976 t/m 1984 tussen hen gewerkt en ik zie daar met genoegen naar terug.

Sinds eind 1984 ben ik full-time werkzaam op de afd. psychiatrie van het Sint Radboud-ziekenhuis, waar ik nieuwe collega's heb ontmoet die mij bij de afronding van dit proefschrift hebben bijgestaan of gesteund. Prof.dr. W.J. Berger – hoofd ad interim – stelde mij in de gelegenheid een gedeelte van mijn werktijd aan dit proefschrift te wijden. Prof.dr. G.J. Zwanikken nam dit over bij zijn benoeming in februari 1985. Door zijn grote interesse voor programma-evaluatie en zijn stimulerende invloed zal deze nu ook op de afdeling Psychiatrie worden ingevoerd. In de stafbarak heb ik tevens veel steun en hulp gehad van de bibliothecaresse Mevr. C. Busser, en de secretaressen Mevr. M. Buchner en Mevr. Th. Verplak.

Tenslotte wil ik stilstaan bij diegenen die een groot aandeel hebben geleverd bij de onderzoekswerkzaamheden zelf: de verwerving en verwerking van de gegevens en de verslaglegging.

– Mevr. W. Rutgers-Greup heeft als projectassistente sinds 1976 de administratie, verzending en verwerking van alle evaluatieformulieren verzorgd. Ook het typen van dit proefschrift is door haar verricht. Haar inzet en accuratesse was onontbeerlijk.

– Ing. H. Wientjes is sinds 1980 als programmeur voor dit project werkzaam. Hij vertaalde al mijn vragen in programma's en leverde zo in totaal vele dozen computer-output en gebruikte hoegenaamd alle SPSSX-procedures.

– Mevr. T. Classens verzorgde de codering van de problemen. Zij stelde een rubriekenlijst samen en zette alle problemen die door de patiënt waren geformuleerd, om in codes. Ook anderszins kon ik steeds op haar een beroep doen.

– Gedurende de 8 jaren dat dit project is uitgevoerd, hebben ook opeenvolgende secretarissen hiervoor diverse administratieve werkzaamheden verricht Mevr J Janssen-Daalen, Mevr Y Heijnen en Mevr M Sloot Zij verzorgden o a de verwerking van de psychologische vragenlijsten Hier moet ook vermeld worden Ir N Cox (U R C) die het programma voor de I D E (M)-procedure heeft ontwikkeld en ook anderszins mij met programmatuur heeft geholpen

Ook de ondersteunende diensten wil ik hier noemen, zoals de afd reprografie, de postkamer, alsmede de Heer W IJsenbrandt, die het subsidiebudget beheerde Het Universitair Rekencentrum verdient een speciaal woord van dank Bij tal van medewerkers, en met name M Manders, kon ik steeds met al mijn vragen en problemen op computer-technisch gebied terecht Gedeelten van het manuscript zijn becommentarieerd door Dr M Lange De Engelse vertaling is verzorgd door de Heer H Tanke en Mevr Drs C Jansen-Nawas

Met inspanning van al deze medewerkers en medewerkers is dit onderzoek voltooid Bij de afronding van het manuscript heb ik veel steun gehad van mijn paranymphen H Reyerman die de tekst op Nederlands heeft nagelezen en zonodig gecorrigeerd, en P Scholte die de lay-out en druktechnische aspecten heeft behartigd In nauwe samenspraak met deze twee vrienden is het eindresultaat van dit proefschrift tot stand gekomen De “finishing touch” werd aangebracht door de Heer J Akkers van Drukkerij MacDonald ten aanzien van de omslag en opmaak van het uiteindelijke boek

Na deze lange rij van personen die hun bijdrage hebben geleverd aan dit proefschrift resteren nog drie mensen mijn vrouw, Maria en mijn zonen Paul en Anko Hun aandeel is zowel actief – Maria was al die jaren een belangrijke gesprekspartner en heeft tenslotte het manuscript nauwgezet op typefouten en andere onvolkomenheden gecorrigeerd, Paul heeft de lokale computerwerkzaamheden ondersteund, – als passief ze hebben mij alle tijd en gelegenheid gegeven om dit project af te ronden en al mijn verhalen en vorderingen aangehoord Ik weet dat ze dat van harte hebben gedaan

Julie allen, mijn hartelijke dank

No Sijben

Inleiding: de Vraagstellingen

De geestelijke gezondheidszorg (g.g.z.) is in hoofdzaak een "non-profit" organisatie. Afgezien van een kleine groep van vrijgevestigden is de g.g.z. ondergebracht in stichtingen die zonder uitzondering tot doel hebben patiënten/cliënten te behandelen zonder winstoogmerk. De geboden hulp is, afgezien van een kleine eigen bijdrage, kosteloos voor de hulpzoeker en wordt uit algemene middelen of door een verzekeringsmaatschappij betaald. Dit is ongetwijfeld een groot goed van onze verzorgingsstaat, maar voor de g.g.z.-organisatie zelf is dit niet een onverdeeld voordeel. Immers, gepaard met de monopoliepositie van g.g.z.-instellingen betekent dit, dat er nauwelijks of geen externe druk is de hulpverlening zo doelmatig en doeltreffend mogelijk te maken. De controle van de overheid, de financier, is voornamelijk of uitsluitend gericht op het budget en de condities waaronder en door wie de hulp gegeven mag worden. De cliënten zelf hebben nauwelijks of geen mogelijkheden elders betaalbare hulp te zoeken als zij ontevreden zijn over de kwaliteit van het gebodene. Zo wordt de g.g.z. een belangrijk sturend mechanisme onthouden zoals dit in de *profit-sector* wel geldt, waar ontevreden klanten naar de concurrent kunnen gaan en slechte bedrijfsresultaten zullen leiden tot ingrijpen van de directie of het bestuur en uiteindelijk eventueel zelfs tot sluiting van het desbetreffende bedrijf of filiaal.

Met deze situatie in de g.g.z. is evenwel niet iedereen tevreden: de cliënten, mogelijk georganiseerd in cliëntenbonden, wensen invloed op beleid en kwaliteit van de zorg. De overheid wenst inzicht in de doelmatigheid en effectiviteit ervan. De derde partij: de werkers in de g.g.z. zullen dit over het algemeen eveneens wel wensen, maar zij voelen zich vaak onmachtig deze effectiviteit aan te tonen; hen ontbreken de mogelijkheden en de middelen. Tevens zullen zij evenwel mogelijke consequenties bij onverhoopte ongunstige resultaten vrezen.

In de jaren zestig ontstaat in de U.S.A. een evaluatieprocedure die erop gericht is deze doelmatigheid en doeltreffendheid aan te tonen met de bedoeling deze zonodig te verbeteren. Deze vorm van "evaluation research" wordt aangeduid met "Program Evaluation". Eind jaren zeventig krijgt deze ook in Nederland toenemende bekendheid onder de naam *Program(ma)-Evaluatie*. Als definitie voor programma-evaluatie kan gelden: *het vaststellen van de waarde van een programma door de effecten ervan te vergelijken met de doelen die voor dit programma zijn gesteld*. Deze evaluatie kenmerkt zich door een sterke praktijkgebonden opzet: het belangrijkste doel is feedback geven over de effectiviteit van een of meer aspecten van het functioneren van een bepaalde instelling: organisatorisch en/of inhoudelijk, op patiënt- c.q. cliëntni-

veau en/of per team, afdeling of instelling als geheel, ten behoeve van het interne management of voor externe instanties.

De twee hoofdmotieven van programma-evaluatie zijn globaal gesproken: *intrinsiek*, de eigen werkers c.q. de directie wensen kennis te hebben van het effect van het functioneren, en *extrinsiek*, de overheid c.q. de financiers wensen geïnformeerd te worden over de besteding van hun gelden of het welzijn van haar burgers. Deze laatste heeft wellicht een belangrijkere stimulans tot programma-evaluatie gegeven dan de intrinsieke motieven.

Dit extrinsieke motief van programma-evaluatie is evenwel niet nieuw. Reeds in 1816 maakte Koning Willem I zich bezorgd over het lot van de zogenaamde krankzinnigen en heeft hij bij herhaling onderzoek laten verrichten naar de omstandigheden van hun verzorging en de effectiviteit van hun behandeling. In hoeverre dit uitmondde in een vorm van programma-evaluatie wordt onderzocht middels een literatuurstudie van de *geschiedenis van het staatstoezicht op krankzinnigen* over de periode 1816-1984.

Het doel en de werkwijze van programma-evaluatie is in eerste instantie descriptief: registreren van onderdelen van het functioneren, de werkwijze en de resultaten. Dit leidt bijgevolg tot de academische vraag - in de letterlijke betekenis van het woord - of programma-evaluatie eigenlijk wel (*wetenschappelijk*) *onderzoek* is of "slechts" een administratieve procedure. Welke plaats neemt programma-evaluatie in binnen het brede scala van sociaal wetenschappelijk onderzoek?

Uit de gegeven definitie van programma-evaluatie zal duidelijk zijn, dat deze zich in vele gedaantes kan voordoen: welke doelstellingen worden geëvalueerd, op welke wijze, met welke middelen, van welke informanten worden de gegevens verkregen. Wat is of kan de rol van de evaluator zijn en hoe wordt zijn informatie gebruikt. Een beschouwing van deze aspecten moet zich voornamelijk baseren op Amerikaanse literatuur, waar sedert 1967 vele tientallen boeken aan dit onderwerp zijn gewijd. Een overzicht van de belangrijkste aspecten van programma-evaluatie om dit *container-begrip* nader te vullen, is zeer gewenst.

Inmiddels zijn of worden ook in Nederland binnen enkele G.G.Z.-instellingen programma-evaluatie projecten uitgevoerd. Hieruit blijkt reeds de verscheidenheid aan vraagstellingen en onderzoeksprocedures. Wanneer programma-evaluatie zich met name richt op het behandelingsresultaat zelf - het moeilijkste te evalueren onderdeel - raakt zij ten volle betrokken in het *operationaliseringsprobleem van de effecten van psychotherapie*: hoe, wanneer en door wie kan worden bepaald wat deze effecten zijn? In tegenstelling tot het meer experimentele psychotherapie-onderzoek waarbij naar statistisch significante verschillen wordt gezocht om dit effect aan te tonen, dient bij programma-evaluatie te worden aangegeven of een

geconstateerd effect voldoende is, of het in overeenstemming is met de vooraf gestelde doelen en zonodig of het in verhouding staat tot de aangewende middelen. Een overzicht van de gangbare operationalisaties dient gegeven te worden, teneinde de keuze van het gebruikte instrumentarium te kunnen motiveren.

Hiermee is het theoretische raamwerk van het hier gepresenteerde onderzoek gegeven. In *deel A, Programma-evaluatie in theorie*, wordt derhalve achtereenvolgens behandeld:

- het staatstoezicht op de krankzinnigenzorg in het licht van hun programma-evaluatieve functie (hoofdstuk 1);
- de plaats van programma-evaluatie in het sociaal wetenschappelijk onderzoek (hoofdstuk 2);
- de praktijk van programma-evaluatie en zijn verschijningsvormen (hoofdstuk 3);
- het meten van psychotherapie-effecten (hoofdstuk 4).

In *deel B, Programma-Evaluatie in praktijk*, wordt een project beschreven, dat gedurende ca. 9 jaar in een kleine kliniek (30 bedden) voor psychotherapie is uitgevoerd. Hierin zijn gegevens verzameld van patiënten die in de periode 1974-1985 ontslagen zijn. Aangezien programma-evaluatie een sterk praktijkgebonden vorm van onderzoek is, dient de *setting* waarin het wordt uitgevoerd goed beschreven te worden om doelstelling en evaluatie-procedures te kunnen toelichten (hoofdstuk 5).

Het *opzetten van een programma-evaluatie-project* is als het aanleggen van een meet- en regelnetwerk, van een nieuw leidingensysteem in een in bedrijf zijnde fabriek. Er moet zorgvuldig rekening worden gehouden met de bestaande situatie bij het invlechten van nieuwe meetprocedures. Hoe dit in dit project is verlopen, wordt beschreven in hoofdstuk 6.

Het uitgevoerde project heeft zich in een viertal fasen voltrokken, waarbij op basis van de ervaringen of de behoefte aan specifieke informatie de meetinstrumenten zijn veranderd. In de periode 1980-1983 is er een vergelijkende studie verricht tussen verschillende vormen van operationalisaties die alle pretenderen (aspecten van) psychotherapie-effecten te meten. Dit arbeidsintensieve onderzoek is mede mogelijk geweest door een subsidie van het Praeventiefonds.

Gedurende de genoemde 9 jaren van de duur van dit programma-evaluatie-project zijn van ca. 900 patiënten gegevens verzameld. Ten eerste betreffen deze de meer objectieve informatie als demografische variabelen, opnameduur, problematiek, hulpverlening voor opname, e.d. Deze informatie schetst de behandelde patiëntengroep en wordt doorgaans in jaarverslagen vermeld. Bij programma-evaluatie zijn deze *registratiegegevens* om meerdere redenen van belang: naast descriptie van de patiëntenpopulatie kan nagegaan worden of deze voldoet aan de doelgroep. Aldus kunnen ze gebruikt worden ter evaluatie van werving en instroom van de te behandelende patiënten conform de doelstellingen. In hoofdstuk 7

worden deze patiëntenkenmerken beschreven en worden de doelstellingen die hiermee verband houden geëvalueerd.

De behandelingsuitkomsten worden in hoofdstuk 8 geëvalueerd. In drie perioden worden *de resultaten van de behandeling* vermeld. Na een retrospectief follow-up onderzoek bij patiënten die in de periode 1974 - 1976 zijn ontslagen, als pilot-studie, volgden twee perioden van permanente evaluatie door de patiënt en zijn hulpverlener bij opname, ontslag en drie follow-up momenten. Deze perioden liepen van november 1976 tot en met 1980, en van 1981 tot medio 1985. Door een verschil in de gebruikte meetinstrumenten worden deze twee perioden apart beschreven.

Vervolgens wordt in hoofdstuk 9 aandacht besteed aan mogelijke *verbanden* tussen deze behandelingsuitkomsten onderling en met patiënt- en therapeutvariabelen. Hiertoe worden de grote hoeveelheid evaluatiegegevens gecomprimeerd tot enkele samengestelde variabelen middels factoranalyse. De registratiegegevens kunnen dan gerelateerd worden aan de behandelingsuitkomsten om mogelijke verbanden op te sporen. Strict genomen overstijgt dit de doelstelling van programma-evaluatie: het evalueren van de behandeling. Het waarom van de al dan niet opgetreden verbetering is een vraag van een andere orde. Het ligt evenwel voor de hand om alle beschikbare informatie met elkaar in verband te brengen, teneinde zo mogelijk ook te kunnen aangeven in welke richting eventuele beleidsveranderingen gezocht kunnen worden. Hiermee is nog geen *verklaring* gegeven voor gevonden relaties.

Een *vergelijkende studie* tussen de diverse meetinstrumenten: gestandaardiseerde psychologische vragenlijsten, beoordelingen van de cliënten van het behandelingsresultaat en diverse doel- c.q. probleemevaluatieprocedures wordt beschreven in hoofdstuk 10. Nagegaan wordt wat hun onderlinge relatie is met betrekking tot het meten van de behandelingseffecten en hoe ze inhoudelijk met elkaar overeenstemmen. Ook de bruikbaarheid van deze meetprocedures wordt beschreven.

Na beëindiging van een dergelijk langlopend programma-evaluatie-project is het gewenst terug te zien op het functioneren van het project zelf teneinde dit eveneens aan een kritische blik te onderwerpen: de *programma-evaluatie* wordt zelf *geëvalueerd* in hoofdstuk 11. Alsmede worden *aanbevelingen* gegeven voor de methodiek, de strategie, de werkwijze en de positie van de uitvoerder van een programma-evaluatie project.

Hiermee is de beschrijving van dit zeer uitgebreide en langlopende evaluatie-project beëindigd. Hoewel programma-evaluatie sterk aan de geëvalueerde instelling gebonden is, levert het ook kennis en ervaring op die ook voor andere (G.)G.Z.-instellingen bruikbaar zijn. Dit geldt met name voor de evaluatiemethodiek en de gebruikte procedures en instrumenten. Daarnaast zal de ervaring van het opzetten en uitvoeren van een dergelijk project, de feedback van de

resultaten en de interactie tussen evaluator en medewerkers van de geëvalueerde instelling aspecten hebben die meer *algemene geldigheid* kunnen bezitten.

Resumerend liggen de vraagstellingen van dit onderzoek op drie niveaus:

- *Descriptie*. Er wordt een beschrijving gegeven van een praktijkvoorbeeld van programma-evaluatie project in een klinische G.G.Z. instelling. In welke setting is dit project uitgevoerd, hoe is het opgezet en welke informatie heeft dit opgeleverd? (hoofdstuk 5, 6, 7 en 8).

- *Vergelijking en verklaring*. Met verschillende instrumenten wordt nagegaan op welke manier de resultaten het beste meetbaar zijn en welke variabelen de effectiviteit en andere aspecten van het functioneren beïnvloeden (hoofdstuk 9, 10).

- *Meta-evaluatie*. Welke conclusies zijn uit een dergelijk project te trekken, hoe bruikbaar is de verzamelde informatie en in hoeverre wordt deze gebruikt. Is programma-evaluatie een nuttige methodiek om de effectiviteit aan te tonen c.q. te helpen verbeteren en aan welke voorwaarden moet deze dan voldoen (hoofdstuk 11)?

Tenslotte worden de belangrijkste resultaten en ervaringen van zowel het theoretische als praktische deel samengevat.

Dit eerste langlopende programma-evaluatie project in Nederland kan het grote en groeiende belang van programma-evaluatie in de (g.)g.z. dienen. Middels programma-evaluatie kan meer worden *omgezien* naar het "*weldoen*" van de G.G.Z. in zijn algemeenheid en van de medewerkers in het bijzonder. Dit zal uiteindelijk de patiënten en cliënten ten goede komen: hun welbevinden is het belangrijkste doel en hoogste *profijt* van de geestelijke gezondheidszorg.

Deel A:

Programma-evaluatie in theorie

Hoofdstuk 1:

Omzien naar Weldoen

Staatstoezicht als Programma-evaluatie

*Men geve in de medische opleiding
een ruime plaats aan het onderwijs
van de geschiedenis der Geneeskunde.
(Kraus, 1933, dissertatiestelling.)*

*En in de opleiding voor psychiaters
aan het onderwijs in de geschiedenis van
de zorg voor de psychiatrische patiënt*

1.1 INLEIDING

"Doe wel en zie niet om" is een welbekend spreekwoord, waarmee aangegeven wordt dat het goede (hulp of andere goede daden) gedaan moet worden zonder op de gevolgen te letten en zonder dank of lof daarvoor te verlangen (Van Dale, 1984). Deze spreuk zou in het vaandel van de (geestelijke) gezondheidszorg geschreven kunnen staan. Van oudsher, althans in de westerse geschiedenis, komt de geestelijke gezondheidszorg voort uit de "edele deugd" van de naastenliefde waarbij zonder winstbejag en liefst met louter altruïstische motieven hulp werd gegeven aan de armen en de anderszins behoeftigen ("bezetenen" of "krankzinnigen") in de samenleving. Deze weldoenersfunctie was aanvankelijk in hoofdzaak voorbehouden aan kerkelijke instanties en welgestelden, zij konden zich dit veroorloven.

In dit hoofdstuk wordt een historisch overzicht gegeven van de ontwikkeling van de zorg voor de psychiatrische patiënt; afhankelijk van liefdadigheid in de middeleeuwen tot een door de overheid bestuurde en gecontroleerde zorg. Met name die controle, *het staatstoezicht op krankzinnigen*, wordt uitgebreid beschreven. In opeenvolgende periodes wordt de groei van het aantal klinische psychiatrische patiënten geschetst en de pogingen van overheidswege om hun behandeling te verbeteren. Uit deze geschiedenis is lering te trekken voor de huidige situatie, of leert de geschiedenis weer dat we niets van de geschiedenis leren?

Stam (1975) meldt in een overzicht over de geschiedenis van de psychiatrie in Nederland dat het eerste zogenaamde "*Dolhuys*" in Nederland gesticht werd in 1425 te Zutphen door Engelbert Kreyninc en de kloosterorde van de broeders en zusters van het "gemene leven". In 1442 en 1461 lieten resp. Reinier van Arkel in Den Bosch en Willem Arntz in Utrecht een fonds na voor de stichting van een instelling voor psychiatrische patiënten. Het Dolhuys in Amsterdam werd in 1561 gesticht met behulp van een gift van 3000 gulden van Hendrik Paulesz. Boelens. Met de stichting van deze huizen probeerden men ongetwijfeld wel te doen, het waren uitingen van naastenliefde, maar naar het welzijn van de patiënten zelf werd verder weinig omgezien. De "razenden" werden doorgaans geboeid of vastgeklonken aan de muur. Het doel van deze opsluiting was meer om de maatschappij voor deze waanzinnigen te beschermen dan de waanzinnigen zelf te helpen. Eenmaal in (geboeide) bewaring gesteld, dienden zij soms zelfs tot publiekelijk vermaak. Tegen betaling konden de gekken bezichtigd worden en zelfs gesard, hetgeen een hoogtepunt was gedurende de zogenaamde "*dolhuyskermis*". Van *psychiatrische zorg* voor de patiënt was geen sprake.

De ambulante psychiatrische patiënt was er gedurende de middeleeuwen soms nog erger aan toe: hij werd vaak opgejaagd van stad naar stad, omdat geen stadsbestuur de kosten van zijn onderhoud wenste te dragen. De enige ambulante "zorg" bestond uit bedevaarten, exorcisme of zogenaamde "water-shock therapie" waarbij patiënten in de Senne werden gegooid. Velen werden het slachtoffer van rondtrekkende charlatans die op gruwelijke wijze de waanzinnigen bevrijdden van "de steen in hun hoofd", waarvan men dacht dat deze de waanzin veroorzaakte (cf. "*De Keisnijder*", Jeroen Bosch). Deze laatste "zorg" spekte uitsluitend de beurs van de charlatan. Voor zover de zinnelozen niet gevaarlijk waren, werden ze dikwijls in de eigen gemeenschap opgenomen of in heiligencentra behandeld. Het dorp Geel in België is bekend om zijn gezinsverpleging. Krankzinnigen die hier ter bedevaart gingen en 9 dagen dienden te blijven om de heilige Dymphna te aanbidden, werden in gezinnen in het dorp opgenomen (Romme, 1984). In hoeverre deze hulp effectief was, verhaalt de historie niet. De naastenliefde van de bevolking van Geel werd alom geprezen, ze "deed wel", zonder om te zien.

Vanaf de 18e eeuw werd de zorg voor de psychiatrische patiënt steeds meer een overheidstaak. Het opsluiten van geesteszieken nam steeds meer toe. De motieven waren aanvankelijk meer economisch en repressief dan humaan. "Zinnelozen werden samen met werkschuwen, bedelaars, leeglopers, havelozen, avonturiers opgesloten in de leeggekomen leprahuizen, omdat arbeid als een ethische waarde werd ervaren en als middel voor het handhaven van de sociale orde. Allen die niet aan de gestelde normen konden of wilden voldoen, werden opgesloten" (Peeters, 1982, p. 15-16) "...Weggestopt achter de coulissen...", (p.13). Deze ongedifferentieerde opsluiting

veranderde rond 1800. De behoefte aan arbeidskrachten maakte een splitsing noodzakelijk van geesteszieken en mensen zonder werk, woon- of verblijfplaats. Tevens ontstond het idee dat de geesteszieken *behandeling* behoeften, dat het milieu therapeutisch moest werken en er werd naar de oorzaken van de stoornissen gezocht. Belangrijke grondleggers van de psychiatrische behandeling waren met name Francis Willis, William Tuke, Philippe Pinel (zie Binneveld, 1985).

1.3 HET BEGIN VAN STAATSBEMOEIENIS: 1800-1840.

Joan van den Honert (1841) beschrijft in zijn boek: *"Geschiedenis en beginselen der Nederlandsche wetgeving betreffende de gestichten voor krankzinnigen, en de wijze hunner opneming in en ontslag uit dezelve"*, hoe de "jonge regering" in Den Haag sedert 1814 geprobeerd heeft de behandeling van de krankzinnigen en het beheer van de inrichtingen te regelen en te controleren.

In de lijn van het in de vorige paragraaf beschreven proces van gedifferentieerde opsluiting bepaalde Koning Willem I bij "Vorstelijk Besluit" van 12 februari 1814, nr. 4, dat krankzinnigen alleen nog mochten worden opgenomen door tussenkomsten van de "Regtbanken van Eersten Aanleg" (art. 1) en "dan nadat het Openbaar Ministerie in zijn conclusien zal zijn gehoord" (art. 2). De bevoegdheid tot opname mocht een jaar duren, steeds met verlenging van een jaar "na verhoor van het Openbaar Ministerie" (art. 3). Hiermee werd bij wet geregeld dat opname en verblijf in een "*verbeterhuis*" slechts onder (juridische) controle kon geschieden.

In 1816 werd door "het Departement van Binnenlandschen Zaken een verslag over het getal, het onderhoud en de bewaring der krankzinnigen aangeboden" (in: Schuurmans Stekhoven, 1922, Rapport van de Staatscommissie d.d. 28 Maart 1827). Kennelijk waren de resultaten van dien aard dat bij vorstelijk besluit van 11 april 1818 werd bepaald, dat nagegaan diende te worden of de bestaande gestichten aan hun doel beantwoordden, nl. "*het genezen van die ongelukkigen*" en indien niet dat ze dan zouden worden *opgeheven* (art. 1) of *verbeterd* (art. 2) (zie afb. 1.1).

De overheid trachtte kennelijk daadwerkelijk om te zien naar het lot van de krankzinnigen en stelde voor hun verblijfplaats te toetsen aan de doelstelling van de aldaar verblijvenden. Voorwaar een vooruitstrevend besluit. De uitvoering van dit besluit werd uiteindelijk opgedragen aan een staatscommissie die bij Koninklijk Besluit van 3 januari 1822 werd ingesteld. Deze commissie hield vervolgens een eerste enquête onder de toen bekende gestichten. Een foto van het oorspronkelijke manuscript van deze eerste enquête naar het aantal krankzinnigen in de Nederlanden, is afgebeeld op pag. 14 en 15 (afb. 1.2)

1°. Dat alle aanwezige gestichten voor krankzinnigen, die be- hoorlijk beantwoorden aan het doel van derzelver stichting, dat is het genezen van die ongelukkigen, zullen blijven bestaan, en daarentegen die welke door hunne ligging, bouwning, of andere omstandigheden, aan het doel niet beantwoorden, zullen worden opgeheven.

2°. Dat de noodige verbeteringen zullen toegebracht worden aan diegene, welke voor zoodanige verbeteringen of voor uitbreiding vatbaar zijn, met in het oog houding dat zoo veel mogelijk in die gestichten een afzonderlijk lokaal zal moeten worden bestemd tot het bewaren van personen die, schoon aan geene dadelijke mis- daad schuldig, wegens wangedrag krachtens de bepalingen der Wet moeten worden opgesloten.

3°. Dat allengskens nieuwe gestichten van dien aard in ge- noegzamen getale en daar waar een geschikt lokaal en andere voordeelige omstandigheden, zoo als de goede koop van aller- lei levensmiddelen, zich zullen vereenigen met de plaatselijke behoeften, zullen worden opgericht;

Afb. 1.1 Koninklijk Besluit van 11 april 1818.

Aldus werd een overzicht verkregen van "het gezamenlijk getal der krankzinnigen". Dat bestond voor de gehele Neder- landen anno 1825 uit 2862 mannen en 2729 vrouwen, in totaal 5591 personen bij een totale bevolking van 5.992.666, hetgeen 0,093% is. 36% Van de mannen en 42% van de vrouwen verbleven in 93 huizen, 46 in de Zuidelijke en 47 in de Noordelijke provincies gelegen. Beperken we ons tot de Noordelijke pro- vincies (ongeveer het huidige Nederland excl. Limburg) dan telde men 1828 krankzinnigen (868 mannen en 960 vrouwen), hetgeen 8.11 per 10.000 zielen was; 707 patiënten (39%) bevonden zich in gestichten. De overigen waren bij partiku- lieren ondergebracht. Van de 47 huizen of gestichten in de Noordelijke provincies zouden slechts 23 bestemd zijn om er krankzinnigen in te verplegen. Van alle opgenomen patiënten was 24% voor eigen rekening opgenomen. De kosten of "beste- dingsgelden" voor hen bedroegen in het Noorden f. 150,- tot f. 200,- per jaar met uitzonderingen van f. 300,- tot f. 600,-.

Naast deze getalsmatige rapportage werd ook een kwalita- tief oordeel over de gestichten gegeven. De conclusie was: *"De Staatsraad Administrateur beschouwt de thans bestaande gestichten, bijna zonder uitzondering, niet anders dan verze- kerde bewaarplaatsen, in welke de krankzinnigen op de eene plaats meer, op de andere met mindere menschlievendheid wor-*

den gevoed, en deze inrigtingen hebben naar zijns inziens in ons Vaderland op verre na dien trap van volkomenheid bereikt, waartoe dezelfde naar het oogmerk hunner bestemming zouden kunnen worden opgevoerd." (Schuurmans Stekhoven, p. 3).

Vervolgens gaf het rapport aanbevelingen voor verbetering. Kennelijk zou men het liefst diverse huizen *"vernietigen of opheffen"* (p. 7). Aanbevolen werd in plaats van de vele kleine gestichten er enkele grotere huizen te bouwen - 5 of 6 dergelijke gestichten voor 500 à 600 krankzinnigen -. Dit zou kosten besparen en de geneeskundige behandeling bevorderen, meende men. Er blijkt op dit en andere punten echter verschil van mening te hebben bestaan tussen de aanbevelingen van de Staatsraad Administrateur P.J. de Beije, en de Staatscommissie. Ook over de verhouding *"geneeslijke"* en *"ongeneeslijke"* krankzinnigen die in een tehuis geplaatst moesten worden, bestond onenigheid. Aanbevolen werd deze twee categorieën patiënten streng te scheiden *"en met aanhouding van afzonderlijke registers voor beide soorten, voor de eigenliefde van den arts, die vordert dat het getal der genezingen beoordeeld kunne worden, alleen in verhouding tot het aantal geneeslijken, dat is opgenomen"*. Terecht merkte de commissie op dat er daarentegen een ander risico zou kunnen bestaan: *"... dezelfde eigenliefde van een arts zou de beweegreden kunnen zijn om bij het bestaan der minste onzekerheid op de genezing den ongelukkigen dadelijk onder het getal der ongeneeslijken te rangschikken; zodoende zou een omgekeerd resultaat der genezingen worden verkregen, waarvan de krankzinnigen het slagtoffer zouden zijn, vermits zij tot de klasse der ongeneeslijken zijnde overgebracht, de middelen van genezing of verwaarloosd, of in het geheel niet zouden worden aangewend"* (p. 13).

Hoe reeds bij opneming geneeslijk en ongeneeslijk gescheiden kon worden, is een raadsel. Uiteindelijk adviseert deze staatscommissie o.a. alle gestichten onder het toezicht van het Gouvernement te plaatsen, een of twee nieuwe gestichten te bouwen en de bestaande gestichten voorlopig ongemoeid te laten. Dit rapport werd 28 maart 1827 uitgebracht, 9 jaar na het desbetreffende Koninklijk Besluit. Ook toen reeds maalden de ambtelijke molens langzaam.

Volgens een nota van de Minister van Binnenlandse Zaken d.d. 25 februari 1829 (no. 116A) bleek deze grote moeite te hebben met de voorstellen van de staatscommissie. Een verplichte opneming in Rijksgestichten had niet zijn instemming; hij achtte de publieke opinie hiertoe nog niet rijp (Van der Esch, I, z.j.). De Raad van State stemde hiermee vervolgens in. Er dienden nieuwe plannen te komen voor het bouwen van kleinere gestichten, er zou wederom een prijsvraag worden uitgeschreven en door drie geneesheren zou een buitenlandse studiereis worden gemaakt, aldus werd op 7 augustus 1830 besloten.

"De laatstgenoemde Kabinetsbrief van 7 augustus 1830 moet nauwelijks in behandeling zijn genomen toen op 25 augustus de Belgische opstand uitbrak, waardoor vele jaren de belangen

Province

1^{ste} Tabel.
Betrueking hebbende tot de Krankz.

1	2	3	4	5	6	7
<p>Welk is het aantal der in deze Provinsie te land te lande zijnde Frankensingen en te over het overige lande van dezelven overen ver- trede als mede of en overde- ten of te Gedeelten der lande, welke niet anders kan een die gene welcke niet anders Provinsie kan een te lande, het te lande, welke te Gedeelten mogten ten lande of overdelede</p>	<p>te Gedeelten en te overdelede welcke mede welcke een die te lande of ten lande ten lande een een Gedeelte of Gedeelten der lande welcke mede kan een die gene welcke niet anders Provinsie kan een te lande, het te lande, welke te Gedeelten mogten ten lande of overdelede</p>	<p>die te lande of ten lande ten lande een een Gedeelte of Gedeelten der lande welcke mede kan een die gene welcke niet anders Provinsie kan een te lande, het te lande, welke te Gedeelten mogten ten lande of overdelede</p>	<p>die te lande of ten lande ten lande een een Gedeelte of Gedeelten der lande welcke mede kan een die gene welcke niet anders Provinsie kan een te lande, het te lande, welke te Gedeelten mogten ten lande of overdelede</p>	<p>die te lande of ten lande ten lande een een Gedeelte of Gedeelten der lande welcke mede kan een die gene welcke niet anders Provinsie kan een te lande, het te lande, welke te Gedeelten mogten ten lande of overdelede</p>	<p>die te lande of ten lande ten lande een een Gedeelte of Gedeelten der lande welcke mede kan een die gene welcke niet anders Provinsie kan een te lande, het te lande, welke te Gedeelten mogten ten lande of overdelede</p>	<p>die te lande of ten lande ten lande een een Gedeelte of Gedeelten der lande welcke mede kan een die gene welcke niet anders Provinsie kan een te lande, het te lande, welke te Gedeelten mogten ten lande of overdelede</p>

Provincia

2^e Tabel

Betrekking hebbende tot de Huizen of Geslachten in welke

1	2	3	4	5	6	7	8
Plaats	Benaming	het Gedeelte of het gedeelte van des Gemeente of van de Gemeente des plaats is of van eenige parochie aangehoort hebben en <u>welke</u> daar en ingescreven hebben?	het Gedeelte van parochie welke is of op teek ingevon naar het Gedeelte van het Gemeente welke worden ingevon en ingeregistreerd hebben en <u>welke</u> daar ingeregistreerd hebben?	het Gedeelte van parochie welke is of op teek ingevon naar het Gedeelte van het Gemeente welke worden ingevon en ingeregistreerd hebben en <u>welke</u> daar ingeregistreerd hebben?	in een der beide gedeelten van en de Gemeente is of ook ook gehoort hebben of andere gedeelten van een of gedeelte van een der beide gedeelten van en <u>welke</u> daar ingeregistreerd hebben?	van een of gedeelte van des Gemeente het gedeelte van des Gemeente het gedeelte van des Gemeente het gedeelte van des Gemeente	of het gedeelte van des Gemeente het gedeelte van des Gemeente het gedeelte van des Gemeente

[illegible]

krankzinnigen werden opgesloten

9					10					11					12				
Beoordeling, doch alleen van de krankzinnigen. Bewoening, aantal inbegrepen van die gene welke van aanwagings opvreesd kunnen staat behouden. Ten geve maakt, die erke en/of verandering van alle toestelke vanover de in het geestelike begrepen worden, begrepen ten behoudens de geestelike veranderingen					Reys van het kostgeld van aan krankzinnigen in de en aanwagings of aanwagings of ge den van pastoor en in het geestelike vergeestelike of geestelike, de beoordeling geestelike krankzinnigen ten de de van de geestelike den werden aanwagings den van geestelike de geestelike van aanwagings van krank zinnigen, de geestelike					Hoe veel de jaar geestelike kosten be gen van onder hand van het geestelike beoordeling van de aanwagings, de aanwagings van geestelike geestelike van onder hand geestelike van onder hand geestelike van onder hand geestelike van onder hand geestelike van onder hand					Aanmerkingen				
Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal				
Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal				
Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal				
Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal				
Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal				
Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal				
Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal				
Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal				
Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal				
Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal				
Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal				
Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal				
Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal				
Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal				
Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal				
Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal				
Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal				
Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal				
Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal				
Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal				
Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal				
Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal				
Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal				
Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal				
Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal				
Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal				
Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal				
Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal				
Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal				
Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal				
Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal				
Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal				
Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal				
Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal				
Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal				
Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal				
Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal				
Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal				
Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal				
Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal				
Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal				
Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal				
Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal				
Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal				
Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal				
Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal					Aanmerkingen, aantal, aantal, aantal, aantal, aantal				

van een betrekkelijk kleine groep van geesteszieke medeburgers op de achtergrond geraakten bij de grotere staatsbelangen " (Van der Esch, I, p. 15). Inmiddels veranderde er niet veel, totdat twee belangrijke personen uit de geschiedenis van de krankzinnigenzorg op het toneel verschenen: C.J. Feith, referendaris van het Armenwezen en Prof.Dr. J.L.C. Schröder van der Kolk. De eerstgenoemde placht de Minister voor Binnenlandse Zaken te begeleiden op zijn inspectietochten in de gestichten. Hij bracht in 1837 middels een nota de ernstige situatie van de krankzinnigen in de gestichten onder de aandacht van de Minister, terwijl in datzelfde voorjaar op 16 maart 1837, Schröder van der Kolk zijn oratie hield, onder de titel: Oratio de debita cura infaustam maniacorum sortem emendandi eosque sanandi in nostra patria nimis neglecta (Redevoering over de verwaarlozing der vereischte zorg ter leniging van het lot der krankzinnigen en ter genezing derzelve in ons vaderland, 1838, Utrecht). Deze rede attendeerde eveneens met klem op de slechte behandeling van de krankzinnigen.

Daarna volgden de gebeurtenissen elkaar snel op. Feith en Schröder van der Kolk overlegden met elkaar; Feith maakte een analyse van de ontwikkeling sinds 1825 en Schröder van der Kolk schreef op verzoek van het ministerie het rapport: "Schets der volstreckte vereischten van een goed gesticht tot genezing van krankzinnigen, en der wijze van beheer van hetzelfde". Deze werden samen met een circulaire van de Minister op 18 mei 1838 naar de Gedeputeerde Staten gestuurd. In deze circulaire werd nogmaals de aandacht gevestigd op de toestand in sommige bewaarplaatsen. Het volgende citaat is hiervoor indicatief: "... Eene plaatselijke inspectie in sommige gestichten, nog in 1837 bewerkstelligd, heeft doen inzien dat er nog bewaarplaatsen zijn, waar zulke ongelukkigen schier moedernaakt in donkere hokken, achter zware grendels opgesloten, een aanwezen voortslepen, waardoor de menschheid gruw" (v.d. Honert, 1841, p. 7).

In deze circulaire verschijnt de eerste evaluatie van de uitkomsten van de verpleging in een Utrechts krankzinnigenhuis dat renoveerde van "een der gebrekkigsten" tot "een Godshuis gelijk". De bijbehorende uitleg van deze kwalificaties sprak boekdelen. Van 1800 tot 1830 was de verhouding van "de verpleegden tot dat der herstelden slechts ... geweest als 24 tot 1,2". Over de periode 1832-1837 is deze verhouding 2,4 tot 1 of 24 tot 10 (zie afb. 1.3).

Hier was voor het eerst sprake van een doelgerichte evaluatie van het welzijn van de patiënten welke tevens als beleidsuitgangspunt diende: renoveren van bestaande huizen was kennelijk een effectieve oplossing (en goedkoper!). Hoe het criterium voor genezing werd vastgesteld is ook hier niet duidelijk.

Uit de circulaire bleek verder, dat er nog steeds gezocht werd naar mogelijkheden om de particuliere tehuizen te kunnen sluiten. Concurreren met overheidsgestichten in de directe nabijheid, was een van de genoemde mogelijkheden. Aan het

Jaar.	Getal der opgenomen krankzinnigen.			Herstelden.			Evenredigheid tus- schen het getal der herstelden en der op- genomenen.	Aanmer- kingen.
	Man- nen.	Vrou- wen.	Geheel bedrag.	Man- nen.	Vrou- wen.	Geheel bedrag.		
1832—1833	17	5	22	9	1	10	0,45=1:2,3	Onder de herstel- den zijn niet be- grepen 8, die min of meer verbeterd, doch niet volku- men genezen aan hunne betrekkin- gen zijn terug ge- geven. Onder de 217 opge- nomen zijn begre- pen de onnoos- len, wezenlozen, en met vallende ziekten of andere ingewortelde kwa- len behoeve en dus ongeneeslijke lij- ders.
1833—1834	32	12	44	14	»	14	0,31=1:3,1	
1834—1835	26	15	41	16	2	18	0,44=1:2,3	
1835—1836	32	21	53	15	5	20	0,38=1:2,5	
1836—1837	35	22	57	20	6	26	0,46=1:2,3	
1832—1837	142	75	217	74	14	88	0,41=1:2,4	

Afb. 1.3 Eerste evaluatie van het behandelingsresultaat
in: Van den Honert (1841) p. 7

slot van dit rondschrijven bevonden zich tabellen "om al dadelijk, toegepast op den toestand van het tegenwoordigen oogenblik, te doen invullen,". En zo werd er in 1838 een tweede landelijke enquête gehouden naar de stand van zaken in de krankzinnigenhuizen. Als de bijlagen in Van den Honert de resultaten zijn van deze enquête - dat blijkt niet uit de tekst -, dan bevonden zich op dat moment 1925 krankzinnigen in de Noordelijke Nederlanden (931 mannen en 994 vrouwen), 825 van hen (43%) bevonden zich in gestichten of huizen, de overigen bij particulieren. In 15 jaar tijd was dit aantal kennelijk slechts weinig gegroeid, maar niet bekend was hoe-
velen er verscholen waren (letterlijk en figuurlijk) onder de particuliere huisvesting.

Al deze inspanningen, met name die van Schröder van der Kolk, leidden tot een wetsontwerp in 1840, dat in 1841 in de Kamer in behandeling werd genomen.

Samenvattend kan de periode van 1800 tot 1841 beschreven worden als goedbedoelde pogingen om het verblijf en de behan-
deling van de krankzinnigen te verbeteren. Diverse registra-
ties en evaluaties waren uitgevoerd die onveranderlijk de
deplorabele toestand van een groot aantal krankzinnigen aan-
toonden. Uit de eerste evaluatie binnen een instelling werd
geconcludeerd, dat nieuwbouw niet noodzakelijk was, omdat
verbetering in huisvesting en behandeling ook resultaten kon
opleveren: de effecten van de behandeling verbeterden aanmer-

kelijk. De noodzakelijk geachte overheidsmaatregelen bleven echter uit, de wettelijke mogelijkheden waren beperkt evenals de financiële middelen. Aan het eerstgenoemde werd rond 1840 met vereende krachten gewerkt, mede onder bezielende invloed van Schröder van der Kolk en Feith. In toenemende mate werd er van staatswege *omgezien naar het weldoen*, voor zover hiervan überhaupt sprake was.

1.4 DE WET VAN 1841.

Op 29 mei 1841 werd de zgn. *Krankzinnigenwet* aangenomen met 41 stemmen tegen 10 (V.d. Honert, 1841). Hiermee poogde men van overheidswege de zo noodzakelijke hervormingen in de huisvesting en behandeling van de krankzinnigen door te voeren. Uit de inleiding op de wet (zie afb. 1.4) wordt de bedoeling ervan reeds duidelijk.

Gelet op de discussie over het wel of niet bouwen van nieuwe inrichtingen is het citaat van Romme (1984, p. 290) des te opmerkelijker. Citeert Van den Honert "... in daartoe meer bijzonder *ingerigte* gestichten" (p. 47); Romme daarentegen, eveneens de wetstekst citerend, schrijft "... in daartoe meer bijzonder *opgerigte* gestichten". Een wonderlijk verschil in citering van dezelfde wet. Het eerste citaat is het juiste, zoals uit afb. 1.4 blijkt.

In deze wet werd in artikel 1 een definitie gegeven voor krankzinnigen. Onder krankzinnigen werden verstaan: "*allen, die van het geheel of gedeeltelijk vrije gebruik van hunne verstandelijke vermogens beroofd zijn.*" In de volgende artikelen werd bepaald, dat er een onderscheid gemaakt moest worden tussen "*geneeskundige gestichten*" en "*bewaarplaatsen*"; de eerstgenoemde werden door de Koning aangewezen (art. 2). De bewaarplaatsen dienden geleidelijk te verdwijnen en er mochten geen nieuwe meer worden gevestigd (art. 3). Met deze artikelen was in feite het doel van deze wet gesteld: een sanering van de tehuizen waar krankzinnigen verbleven. Maar men ging verder. Zo vermeldde artikel 5:

artikel 5

Er zal op last des Konings een plaatselijk onderzoek geschieden naar de wijze van bewaren en verzorgen der krankzinnigen in de bestaande gestichten; - degene, in welke die wijze bevonden wordt tot verergering van hunnen toestand aanleiding te geven, zullen, op last des Konings, door het administratief gezag worden opgeheven en gesloten, wanneer niet, binnen een daartoe te stellen termijn, zoodanige verbeteringen daaraan worden toegebracht, waardoor die aanleiding ophoudt. Op dezelfde wijze, zullen gesloten worden de gestichten, omtrent welke de bij het vorig artikel bedoelde kennisgeving niet zal zijn gedaan,

STAATSBLAD

VAN

HET KONINGRIJK DER NEDERLANDEN.

(N^o. 20.) *WET van den 29sten Mei 1841,
houdende bepalingen omtrent de
gestichten voor krankzinnigen, en
de wijze hunner opneming in en
ontslag uit dezelve.*

WIJ WILLEM II, BIJ DE GRATIE GODS, KONING,
DER NEDERLANDEN, PRINS VAN ORANJE - NASSAU,
GROOT - HERTOG VAN LUXEMBURG, ENZ., ENZ., ENZ.

Allen die deze zullen zien of hooren lezen, salut!
doen te weten:

Alzoo Wij hebben in overweging genomen, dat de voortgang der wetenschappen, en de, zoo in Nederland als elders, verkregene ondervinding, de overtuiging hebben verwekt, dat de herstelling der krankzinnigen aanmerkelijk bevorderd wordt, door eene spoedige geneeskundige behandeling, in daartoe meer bijzonder ingerigte gestichten;

En willende de vereischte voorzieningen daarstellen om de bereiking van het heilzaam oogmerk van zoodanige gestichten, in onderscheiding van andere, minder aan het voorschreven doel dienstbare inrigtingen, te vereenigen met de verzekering der goede behandeling en behoorlijke waarneming der zakelijke belangen van

Afb. 1.4 Inleiding op de Wet van 29 mei 1841

of die zonder toestemming des Konings mogten worden opgerigt.

(Van den Honert, p. 97).

Dit artikel zou een *evaluatie-artikel* genoemd kunnen worden: nagegaan wordt of de doelstelling van de inrichtingen wordt gerealiseerd. Indien het tegenovergestelde effect bereikt wordt - verergering van de toestand van de opgenomen patiënt - volgt een drastische sanctie: sluiting van de inrichting. In artikel 6 werd bepaald, dat dan eventuele fondsen die bij opheffing vrijkomen ten goede zullen komen aan de verzorging van hun patiënten. Artikel 7 vermeldde dat naast het permanente toezicht van plaatselijke besturen, alle gestichten onder het oppertoezicht van de regering vielen. Als een gesticht niet meer aan zijn bestemming voldeed, kon het conform artikel 5 worden gesloten. De feitelijke toezichthouders op een goede naleving van deze wet waren de officieren bij de arrondissementsrechtbanken samen met een geneeskundige. Zij dienden *"... op onbepaalde tijdstippen, doch telken drie maanden tenminste eenmaal, al die gestichten voor krankzinnigen in hun ressort te bezoeken, om zich te verzekeren, dat niemand wederrechtelijk in dezelve geplaatst is noch teruggehouden wordt, en dat de krankzinnigen behoorlijk worden behandeld"* (art. 9). Hiermee is de essentie van deze wet weergegeven. De volgende artikelen zijn uitvoerende bepalingen betreffende opname (art. 10 t/m 22), ontslag (art. 23 t/m 27) en overige zaken (art. 28 t/m 30).

Met andere ogen bekeken kan hier gesproken worden van een vorm van (*programma-*)*evaluatie*: bekend is dat er ernstige misstanden zijn in de verzorging van de krankzinnigen. Er worden nieuwe doelstellingen geformuleerd, deze worden periodiek geëvalueerd, teruggekoppeld en bij (herhaaldelijk) niet voldoen aan de doelstellingen staan zelfs de sancties vast. Aldus kan deze wet in zijn bedoeling beschouwd worden als een eerste vorm van zgn. externe programma-evaluatie (zie ook hoofdstuk 3.3.2).

Bij Koninklijk Besluit van 5 oktober 1841 werd een landelijk register ingevoerd dat voor iedere opgenomene moest worden aangelegd (zie afb. 1.5, p.22); registratie is ook nu nog de eerste stap in een evaluatie-proces. Tevens werd de controlerende taak nader uitgewerkt in artikel 5 en 6.

Artikel 5.

In den loop van ieder jaar zullen, naar aanleiding van het zevende artikel der voormelde Wet, eenige der bestaande geneeskundige gestichten en bewaarplaatsen, jaarlijks door Onzen Minister van Binnenlandschen Zaken aan te wijzen, worden onderzocht door eenen ambtenaar en eenen geneesheer, daartoe door Onzen genoemden Minister te benoemen. De aanwijzing der ieder jaar te onderzoeken gestichten zal in dier voege plaats hebben, dat ieder der bestaande gestichten uiterlijk telken drie jaren eenmaal daarin begrepen zijn.

Artikel 6.

Het bij het vorig artikel bedoelde onderzoek zal zich uitstrekken tot al hetgeen nopens de inrigting der gestichten, de wijze van besturen derzelve en van het behandelen der daarin geplaatste personen, moet bekend zijn, om de verzekering te erlangen, of en in hoeverre zij aan derzelve bestemming beantwoorden.

Van de bevinding daaromtrent in de telken jaren bezochte gestichten, zal een omstandig en vergelijkend verslag aan Onzen Minister voornoemd worden uitgebragt.

(Van den Honert, p. 250).

De in artikel 5 genoemde ambtenaar en geneesheer werden op 1 maart 1842 benoemd; het wekt geen verbazing dat dit Feith en Schröder van der Kolk zijn. Zij waren bijgevolg de eerste *"Inspecteurs der gestichten voor krankzinnigen"*, hun taak was om te zien naar het *"weldoen"* aan en het *welzijn* van de krankzinnigen. Het feit dat dit een part-time functie was en de transportmiddelen beperkt waren - pas sinds 1870 was er een treinverbinding Den Haag-Utrecht - maakte dat zij hun controlerende taak weinig optimaal konden verrichten.

1.5 HET STAATSTOEZICHT VAN 1842 TOT 1884.

Ondanks de aan het slot van de vorige paragraaf genoemde beperkingen, gingen de twee inspecteurs voortvarend van start. Reeds in oktober 1842 brachten zij verslag uit van hun inspectie van alle 24 niet erkende gestichten - en passant nog 4 onbekende op hun reis tegenkomend. Dit tweede landelijke vergelijkende onderzoek - het eerste was de schriftelijke enquête van de staatscommissie in 1822 - gaf volgens de rapporteurs een schrikbarend beeld van de stand van zaken in de bezochte inrichtingen. Hun slotconclusie luidde: het onderzoek had *"tot de ongelukkige zekerheid geleid dat buiten de reeds als geneeskundig erkende gestichten geen enkel dier gestichten in ons Vaderland aanwezig is, hetwelk niet in mindere of meerdere mate groote gebreken in dit opzigt in zich vereenigt ..."* en bijgevolg *"... te verklaren dat wij geene vrijheid vinden om voor te stellen dat een eenig der door ons thans bezochte gestichten, in den staat in welke wij die bevonden hebben, tot geneeskundig verklaard zoude worden ..."* en tot slot *"... dat het ... niet zonder de treurigste herinneringen en met een pijnlijk gevoel (was) dat wij vele dier gestichten als ware hollen des ongeluks hebben leren kennen"* (Vijselaar, 1983, p. 652). Mochten de inspecteurs of anderen nu al verwacht hebben dat de in de wet genoemde sancties i.c. sluiting van deze gestichten uitgevoerd zouden worden, dan zullen zij in deze teleurgesteld zijn geworden. Het ontbrak de overheid aan middelen al deze gestichten te vervangen door

afb. 1.5 Eerste register volgens Koninklijk Besluit van 5 oktober 1841

REGISTER, gehouden ter volgoening aan de bepalingen der wet van den 29sten Mei 1841
(*Staatsblad* n°. 20), voor het geneeskundig Gesticht (of de bewaarplaats) voor Krankzinnigen,
genaamd te

Volgenummer.	Naam, voornamen, ouderdom, woonplaats van dengene, die de magtiging tot opening heeft gevraagd, of aandeeling van het Openbaar Ministerie door hetwelk die magtiging ambtshalve is getequeerd.	Omschrijving of afschrift der stukken, welke zoo bij de aanvraag tot opening, als later, ten gevolge van art. 18 en 22 der in het hoofd van dit register vermelde wet, zijn overgelegd.	Dagteekeningen van de magtigingen tot voorloopige en definitieve opening van die tot verlenging van het gesticht.	Dagteekeningen van de kennisgeving van het gesticht aan den Officier van Justitie.	Dagteekeningen op de termijnen, voor welke de magtigingen, voor het verblijf in het geesthuis, verstrekt zijn.	Verklaaring van den geneesheer van het geesthuis (of van het den toestand in welken hij, die het geesthuis verlaat, op dat de bepalingen van de wet betrekking heeft, met de kennisgevingen van hoofd van dit register het ontslag aan den Officier van Justitie.	Dagteekening van het verlaten van het geesthuis; de oorzaken van iederen opgenoemen, en de bepalingen van de wet, met betrekking tot de kennisgevingen van hoofd van dit register het ontslag aan den Officier van Justitie.	Vermelding van het ten opzichte van iederen opgenoemen, en de bepalingen van de wet, met betrekking tot de kennisgevingen van hoofd van dit register het ontslag aan den Officier van Justitie.

Behoort bij Koninklijk besluit van den 5den October 1841,
n°. 64 (*Staatsblad* n°. 41).

Ons bekend,
De Minister van Binnenlandse Zaken,
SCHIMMELPENINCK VAN DER OIJDE,
De Minister van Justitie,
VAN M'ANEN.

nieuwe. Daarnaast hadden de inspecteurs geen andere bevoegdheid dan de status quo te signaleren en te rapporteren. Ook in dat opzicht kwam hun functie overeen met die van programma-evaluatoren (zie hfst. 3.5). Zij moeten zich hierbij zeer machteloos gevoeld hebben, het lot van de patiënten trokken zij zich zeer aan.

Inmiddels continueerden de beide inspecteurs hun bezoeken aan de inrichtingen en bewaarplaatsen. In het Historisch Archief te 's-Gravenhage zijn zelfs de oorspronkelijke aantekeningen van iedere inspectietocht over de periode 1838 tot 1859 bewaard. Hierin is te lezen dat zij per inrichting een kort verslag maakten met alle sterke en zwakke punten en adviezen die zij gaven. Bij een volgend bezoek werd nagegaan of deze adviezen waren opgevolgd. Van deze inspectietochten maakten zij een of meer jaarlijkse verslagen. Het eerste verscheen in 1848 over de periode 1844-1846 (Feith en Schröder van der Kolk, 1848). Deze verslagen openden alle met het hoofdstuk: *De werking van de Wet*, waarin gedetailleerd werd aangegeven op welke onderdelen de situatie in een of meer gestichten nog niet aan de wet voldeed. In dit eerste jaarverslag werd ook een vergelijking per gesticht gegeven over de situatie van 1838 en die van 1846, waarin grote kwalitatieve veranderingen werden beschreven.

Ook werd getracht de behandeling te evalueren door bij de uitgeschrevenen een opsplitsing te maken in: overleden, achteraf niet krankzinnig, niet verbeterd, verbeterd en hersteld (p. 34 e.v.). De achteraf-niet-krankzinnigen werden later in de tabellen van het aantal opgenomen afgetrokken (zie jaarverslag 1852). Er werd een *evaluatie-index* berekend door het aantal herstelden te relateren aan het aantal opgenomen. De resultaten per jaar staan in tabel 1.1 vermeld.

Tabel 1.1 Aantallen opgenomen en ontslagen patiënten over de periode 1844-1846.

jaar	----- ontslagen-----				herstel	
	opn.	n. verb.	verb.	hersteld	overl.	ratio
1844	412	13	28	140	134	0,34
1845	383	22	33	125	124	0,33
1846	416	25	36	147	154	0,35
totaal	1211	60	97	412	412	0,34

Voor de bewaarplaatsen werd een dergelijke berekening onjuist gevonden, mede vanwege de kleinere aantallen. Wel werden deze Nederlandse resultaten vergeleken met buitenlandse cijfers. Zo werd voor een achttal Engelse gestichten een herstelcoëfficiënt van 0,37 gemeld. De inspecteurs conclu-

deerden dan ook, dat Nederland zich met de zo geprezen buitenlandse situatie mocht meten (p. 37). Er werd ook een overzicht van deze coëfficiënten per gesticht gepresenteerd. Deze varieerden van 0,165 tot 0,506. Het heeft Schröder van der Kolk kennelijk gestoken dat "zijn" Utrechtse inrichting niet op de eerste plaats stond maar 's-Gravenhage, gezien de beschrijving van de lijstaanvoerder. Uiteindelijk werd gemeld dat uit deze cijfers geen conclusies getrokken konden worden, maar dat eerst over meerdere jaren gegevens verzameld moesten worden. Nadere analyse van deze cijfers leidde tot de conclusie dat er geen verschil is in het verbeteringsquotiënt tussen mannen en vrouwen. Wel bleek dat hoe eerder een patiënt werd opgenomen na de start van de ziekte, des te sneller het herstel intrad. Ook was het herstelquotiënt bij de kortopgenomenen veel hoger dan bij de langer verblijvenden.

De jaarverslagen werden allengs gedetailleerder: naast een getalsmatige beschrijving van aantallen opgenomen en ontslagen patiënten - zo mogelijk uitgesplitst naar diverse persoonskenmerken - werd een kwalitatieve beschrijving gegeven van de behandeling, de huisvesting, het personeel en de overige voorzieningen van een bepaalde inrichting. Later werd ook een beschrijving opgenomen van o.a. de kwaliteit van het drinkwater en de grond waarop de opstallen gebouwd waren. Of al deze informatie relevant was voor het beleid is niet (direct) aantoonbaar. Voor het hier gepresenteerde overzicht zijn de zgn. doorschrijftabellen, welke sinds 1848 in de verslagen zijn opgenomen, van belang. Deze geven een sequentiële beeld van aantal opnamen en ontslagen patiënten, waarbij ook is aangegeven of deze patiënten bij ontslag wel of niet waren verbeterd. Hoewel niet duidelijk is volgens welk criterium deze verbetering werd gemeten, kan dit toch gezien worden als een *permanente evaluatie* van het behandelingsresultaat van de klinisch psychiatrische patiënt.

Hoewel de veranderingen zoals door de wet bedoeld, zich slechts heel geleidelijk voltrokken, bleken 10 jaar na deze wet toch belangrijke wijzigingen in aantal en kwaliteit van de inrichtingen te hebben plaatsgevonden. In 1842, bij de eerste landelijke inspectie, werden 28 gestichten bezocht waarvan er drie reeds de erkenning "*geneeskundig inrichting*" verworven hadden (te Utrecht, Deventer en Zutphen). Tien jaar later, in 1851, bestaan er nog 18 gestichten: 12 geneeskundige inrichtingen en 6 bewaarplaatsen. Deze veranderingen deden zich voor door een vorm van natuurlijk verloop: diverse gestichten sloten zelf hun deuren, uit vrees dat dit t.z.t. van staatswege zou kunnen gebeuren. Of zij waren gedwongen te sluiten bij gebrek aan patiënten, omdat familieleden hun krankzinnige verwanten niet in een ontoereikende "bewaarplaats" wensten te laten opnemen. Zo ontstond er plaatsgebrek en aangezien het provinciale bestuur diende te zorgen voor voldoende opnamecapaciteit, werden er nieuwe inrichtingen gebouwd die van meet af aan aan de vereisten van de wet voldeden. Dit gebeurde in Maastricht, Bloemendaal en Franeker.

Na 10 jaar kan zeker gesproken worden van een geslaagde verandering in de zorg voor de krankzinnigen, zowel kwalitatief als kwantitatief. Een treffend kwalitatief voorbeeld van de bereikte veranderingen geeft het verslag over 1851-1853. De situatie van de krankzinnigen vóór 1838 werd aldus beschreven: *"veelal in de ellendigste verblijven als wilde dieren opgesloten, soms jarenlang van het zonlicht beroofd; aan de ruwheid van bedienden overgelaten, sleepten zij hun ellendig leven in werkeloosheid, meestal jammerlijk verwaarloosd voort, zonder noemenswaardige geneeskundige verzorging; ... terwijl men niet anders dacht dan wel te doen, ..."* Over de verslagperiode, 15 jaar later, werd geoordeeld dat *"deze verblijfplaatsen van jammer en der hoogste menselijke ellende, in verblijven van dankbaarheid voor genoten weldaden zijn 'herschappen' ..., alle vroegere sporen van geweld zijn verdwenen; alle boeien en ketenen gestaakt; alle barbaarsche behandeling heeft opgehouden ... toedienen van dienstig uitgekozene en welberekende middelen tot herstel, ... waar vroeger de lijders verscholen leefden, zijn zij thans allen, niet slechts onder het wakende oog en de bescherming van de Regtbanken en de Hooge Regeering geplaatst, en is alle vrees voor onwettige opsluiting verdwenen, maar ondervinden zij ook gestadig de gezegende vruchten van de ijverige zorgen van de bestuurders ..."* (Van der Esch, I z.j., p. 25). De periodieke inspectietochten en de verslagen en adviezen van de inspecteurs hebben hierop zeker grote invloed gehad. Deze eerste *programma-evaluatoren* in de klinisch-psychiatrische zorg hebben zeker succes gehad: ondanks het feit dat het hen aan machtsmiddelen ontbrak, leidde hun voortdurende rapportage aan de minister en de adviezen die zij gaven aan de directies van de inrichtingen, uiteindelijk toch tot het beoogde doel.

Naast de functie van controleur en rapporteur bouwden de inspecteurs een grote psychiatrische expertise op. Zij kregen een overzicht van alle opgenomen patiënten, hun ziektebeelden en hun behandeling. In de beginjaren sprak Schröder van der Kolk met elke nieuwe patiënt. Het ligt dan ook voor de hand dat het eerste Nederlandstalige handboek voor de psychiatrie door hem geschreven is (Schröder van der Kolk, 1863). Het verdere verloop van de periode van 1845 tot 1885 kenmerkt zich dan hoofdzakelijk als een consolidering van het bereikte en een gestage terugdringing het aantal bewaarplaatsen. In tabel 1.2 wordt een overzicht gegeven van het aantal gestichten gesplitst naar inrichting en bewaarplaats met het aantal opgenomen en ontslagen patiënten per 10 jaar. Als bron hiervoor heeft het titanenwerk van Schuurmans Stekhoven (1922) gediend. Deze inspecteur, van 1901-1934, heeft een overzicht samengesteld van honderd jaar krankzinnigenwezen in Nederland (1813-1914) in tabellen, grafieken, tekeningen en dergelijke. Een groot gedeelte is gebaseerd op de verslagen van de inspecteurs sinds 1844. De onderverdeling: niet verbeterd, verbeterd en hersteld, heeft hij teruggebracht tot een tweedeling: niet verbeterd en verbeterd werden samengenomen.

Tabel 1.2 Overzicht van opnamen en ontslagen over de periode 1845-1885 uitgesplitst naar type instelling, in 10 jaar-periodes.

jaar	aard	N	aanw.	opg.	----ontslagen----			
					niet hersteld	wel overl.	accres	
1845	g.i.	11	934	383	55	125	124	79
	bew.	11	138	10	33	6	6	-36
1855	g.i.	12	1278	627	62	233	232	100
	bew.	5	66	7	-	3	3	1
1865	g.i.	12	2730	912	86	345	334	147
	bew.	2	62	1	-	-	2	-1
1875	g.i.	13	3753	1028	136	386	403	103
	bew.	2	59	2	-	-	3	-1
1885	g.i.	16	4843	1266	135	386	459	286
	bew.	1	5	-	-	-	-	-

(naar: Schuurmans Stekhoven, 1922, tabel V)

Verklaring: g.i.= geneeskundige inrichting;
 bew.= bewaarplaats; aanw.= aanwezig op 1 januari;
 opg.= opgenomen; overl.= overleden;
 accres = jaarlijkse toename.

Uit dit overzicht blijkt, dat het aantal bewaarplaatsen snel verminderde: het aantal opnamen hierin was in 1845 nog slechts 10 en nam daarna nog verder af. Vanwege de daar reeds jaren verblijvende patiënten werden ze niet gesloten. In 1886 werd de laatste bewaarplaats, het Cellebroederenhuis te Nijmegen, opgeheven. De 4 laatste patiënten werden "niet hersteld" ontslagen, waarschijnlijk overgeplaatst, zoals dit met de meeste niet hersteld ontslagenen gebeurde.

Uit het overzicht van tabel 1.2 blijkt verder, dat in de periode van 1845 tot 1885 een toename van de aanwezige krankzinnigen van 934 tot 4843 personen, ofwel ruim het vijfvoudige, plaatsvond. Gerelateerd aan de totale bevolking steeg het aantal opnamen van ca. 3,5 naar 11,5 per 10.000 inwoners. Het aantal opnames per jaar steeg met 330%. De jaarlijkse toename (accres) bedroeg ca. 100 patiënten. Het herstelquotiënt klom geleidelijk van .33 tot .38. In de opeenvolgende jaarverslagen worden de drie quotiënten: overledenen, niet herstelden en hersteld ontslagenen steeds weergegeven. Het quotiënt voor overledenen is doorgaans hoger (ca. .40) dan van de herstelden (.35). Dat van de niet-hersteld-ontslagenen bedroeg steeds ca. .25.

Hiermee kan een belangrijke periode in de geschiedenis van de krankzinnigenzorg in Nederland worden afgesloten: de wet van 1841 heeft onmiskenbaar zijn nut bewezen en de inspanningen van de inspecteurs, de eerste programma-evaluatoren in de geestelijke gezondheidszorg hebben dit in grote mate bewerkstelligd. In het jaarverslag over 1882-1884 (1891) werd de

wet van 1841 geëvalueerd en de conclusie luidde: *het doel is bereikt*. Er werden evenwel ook leemten geconstateerd, zodat voorbereidingen getroffen werden voor een nieuwe wet.

1.6 DE PERIODE 1885-1936.

In de wet van 1841 werden tekortkomingen geconstateerd. De belangrijkste waren volgens het jaarverslag van 1882-1884:

- de wet had alleen betrekking op krankzinnigen die in gestichten verbleven;
- de scheiding tussen geneeskundige inrichting en bewaarplaats is ongegrond. Inmiddels was dit onderscheid overigens achterhaald;
- er waren geen voorwaarden genoemd waaraan een erkende inrichting behoorde te voldoen;
- onduidelijk was wiens plicht het was om voor voldoende opvang voor krankzinnigen te zorgen. Art. 8 van de wet was bedoeld om dit te regelen, maar door zijn te vage formulering werkte dit in de praktijk niet;
- een grote hoeveelheid formaliteiten die verricht moesten worden, werkte zeer vertragend bij een opname;
- de geldigheidsduur van een machtiging of geneeskundige verklaring was te lang;
- onduidelijk was wie deze machtigingen en verklaringen mocht verlengen;
- ontslag uit een inrichting kon alleen bij goedkeuring van de geneesheer.

Deze leemten waren reeds veel eerder geconstateerd. Met name het onderscheid tussen geneeskundige inrichting en bewaarplaats heeft tot veel discussie geleid. Reeds vóór 1841 had Schneevogt dit onderscheid onjuist gevonden. Overbevolking van de gestichten, met name door het toenemende aantal onder curatele gestelden die zich in een "gedurige staat" van krankzinnigheid bevonden, welke staat volgens art. 20 van de wet na 3 jaar machtiging intrad, was een druk op verandering van de wet. Door deze verblijfspatiënten was nl. de opnamecapaciteit voor acute gevallen zeer beperkt. Dit leidde anno 1854 zelfs weer tot het idee *bewaarplaatsen nieuwe stijl* voor de chronische patiënten op te richten (Van Esch III, p. 67).

De inspecteurs hebben herhaaldelijk verzoeken tot wetswijzigingen ingediend, maar zonder enig resultaat. Zij legden zich kennelijk niet meer bij hun louter descriptieve taak, en wensten hun bevindingen om te zetten in concrete beleidsaanbevelingen. In 1859 was zelfs door Feith en Schröder van der Kolk een ontwerp voor een nieuwe krankzinnigenwet opgesteld. In 1864 dienden Feith en Schneevogt een tweede ontwerp in. De inspecteurs Ramaer en Van Capelle begonnen in 1876 aan een derde ontwerp, dat in 1881 aan de Staten Generaal werd aangeboden. Dit diende als voorontwerp en werd na wijzigingen in 1884 aangenomen" (Van Esch III, p. 68). De nieuwe wet van 27

april 1884 getiteld: *Wet tot regeling van het Staatstoezicht op Krankzinnigen en Krankzinnigengestichten*, trad op 1 oktober van dat jaar in werking. De belangrijkste artikelen waren:

- art. 1, dat het staatstoezicht (cf. de titel van de wet) regelde door o.a. de twee inspecteurs in full-time dienstverband aan te stellen;
- art. 2, waarbij de werking van de wet zich uitstreckte tot alle krankzinnigen, dus ook degenen die buiten de inrichtingen verbleven, behalve de thuiswonenden;
- art. 3 verplichtte dat van alle krankzinnigen aangifte moest worden gedaan;
- art. 4 bepaalde, dat er van alle krankzinnigen, ook die in particuliere woningen, register gehouden moest worden. Een voorbeeldregister werd bij de wet gepresenteerd
- art. 7 e.v. gaven de eisen aan op grond waarvan een vergunning tot verpleging van krankzinnigen afgegeven zou worden. Bij Koninklijk Besluit zou per gesticht het maximale aantal patiënten en het minimale aantal geneesheren worden vastgesteld;
- art. 11 behandelde de taak van de provinciale besturen om voor voldoende opvang en plaatsing van krankzinnigen te zorgen;
- art. 14 regelde de mogelijkheid tot een snellere opnameprocedure: de inbewaringstelling (I.B.S.), welke door de burgemeester van de gemeente kon worden afgegeven voor een duur van 8 dagen;
- art. 28 regelde de ontslagprocedure.

Hiermee was getracht eerdergenoemde leemten op te vullen. Het oordeel over deze wet die blijkens de voorgeschiedenis een aaneenschakeling was van compromissen en afgezwakte maatregelen, is weinig positief. Van der Esch zegt: "*De Krankzinnigenwet van 1884 behoort tot de krakende wagens die het langst duren*" (Van der Esch, III, p. 70). De opeenvolgende jaarverslagen gaven voortdurend aan hoezeer deze wagen "kraakt". Uit het 17e jaarverslag over 1885-1889 blijkt, dat de meldingsplicht van alle krankzinnigen niet werkte. Er bestond grote onduidelijkheid wie als krankzinnig moest worden beschouwd. Er werden van 1 oktober 1884 tot 1 januari 1888 794 krankzinnigen aangemeld, terwijl de inspecteurs er hiervan maar 515 als terecht oordeelden.

In het jaarverslag over 1897, 1898 en 1899 wordt gemeld dat men nog steeds grote problemen had met artikel 3. Er was nl. een nieuwe categorie van patiënten ontstaan, zgn. "zenuwlijders" die in aparte inrichtingen werden verpleegd. De bestuurders van deze tehuizen meenden dat hun patiënten niet onder de wet vielen, ze waren immers niet krankzinnig. Opgemerkt moet worden, dat in de wet van 1884 geen omschrijving meer gegeven was van krankzinnigheid, zoals in 1841. Of het opnemen van een definitie dit probleem had kunnen voorkomen, valt zeer te betwijfelen.

Een essentiële tekortkoming van de wet, die immers het staatstoezicht beoogde te regelen, is het gebrek aan bevoegdheden van de inspecteurs. De inspectie was machteloos tegen een onwillig bestuur, zoals in 1890 blijkt uit een conflict met regenten van het krankzinnigengesticht te 's-Gravenhage. Dit gesticht kende vele gebreken, herhaaldelijk geconstateerd door de inspecteur, maar adviezen tot verbetering werden niet opgevolgd. Sanctiemogelijkheden waren de inspectie onthouden door de Kamer, uit vrees voor een te grote inperking van de bestuurlijke autonomie van de gestichten (Vijselaar, 1983).

In de jaarverslagen werden sinds 1844 steeds in de zgn. doorschrijftabel de aantallen opnamen, ontslagenen (hersteld of niet hersteld) en overledenen vermeld, alsmede het eerder genoemde herstelquotiënt. Schreven Feith en Schröder van der Kolk in het eerste jaarverslag (1848) dat pas conclusies getrokken zouden mogen worden als er meer jaren vervolg zouden zijn, in 1901 vermeldde Ruysch en Van AnDEL: *"Wanneer men aandachtig de hier weergegeven cijfers nagaat, blijven enige schommelingen niet onopgemerkt, zonder ons evenwel te verliezen tot het maken van waarschijnlijk praemature gevolgtrekkingen"* (p. 369). Een wonderlijk commentaar bij een getalsmatig overzicht van 56 jaar krankzinnigenzorg! Als hieruit dan nog geen gevolgtrekkingen te maken zijn, dan zal dit wellicht nooit kunnen. Hierop zal aan het slot van dit hoofdstuk nog worden ingegaan. In het jaarverslag over 1900, 1901 en 1902 (Schuurmans Stekhoven en Van Deventer, 1907) wordt een korte samenvatting gegeven van alle opnamen sedert 1844: 40% is overleden, 24% is niet hersteld ontslagen en 36% is hersteld ontslagen. Deze gemiddelde cijfers wijken nauwelijks af van de eerste cijfers van Feith en Schröder van der Kolk in 1848.

Wat met name opvalt bij de eerste verslagen van Schuurmans Stekhoven en Van Deventer is hun onhanteerbare omvang: het 19e jaarverslag (1900-1902) bevat 556 pagina's, en nr. 20 (1903-1908) zelfs 720 pagina's plus 22 tekeningen, maar dit behandelde 6 verslagjaren. Nummer 21 over 1909-1911 besloeg weer 602 pagina's en 9 tekeningen. Een dergelijke omvangrijke verslaglegging kan nauwelijks bevorderlijk geweest zijn voor mogelijke invloed op het beleid. Er is ook inhoudelijk verschil zichtbaar: het percentage hersteld ontslagen patiënten daalt, over 1903-1908 wordt een gemiddeld herstel van 33% gegeven. In het 21e jaarverslag wordt alleen gemeld, dat het herstel dalende is. In dit en volgende jaarverslagen worden geen herstelpercentages meer gegeven. Worden deze uitgerekend aan de hand van tabel V van Schuurmans Stekhoven, dan blijkt een geleidelijke daling van 35,3% in 1900 tot 29,9% in 1914 met een laagste percentage in 1909 van 27,2%. Het criterium dat volgens de wet het staatstoezicht diende te behartigen - het herstel van de krankzinnigen - wordt op het moment dat blijkt dat dit criterium minder gerealiseerd wordt, weggelaten. Een zeer opmerkelijk verschijnsel, maar niet geheel onbekend op het terrein van de presentatie van evaluatie-resultaten.

Het 25e jaarverslag handelt over 15 jaar: 1915-1929 (Schuurmans Stekhoven, 1932). Voor deze grote periode worden diverse redenen opgevoerd: de Eerste Wereldoorlog, het overlijden van inspecteur Van Deventer (1916) en de herindeling van de inspecties. Uit een overzichtstabel over 1849 tot 1929 blijkt, dat het aantal opnemingen is gestegen van 1.187 naar 21.318 per jaar ofwel van 3,9 naar 27,2 per 10.000 inwoners. Het herstelpercentage daalt van 37% tot 31,1% in 1929 (zie ook tabel 1.3).

Tabel 1.3 Overzicht van de opnamen en ontslagen over de periode 1886 - 1936 in intervallen van 10 jaar

jaar	aanw.	opgen.	-----ontslagen-----			accres
			niet her.	herst.	overl.	
1886	5129	1255	125	360	533	236
1896	7109	1637	210	592	619	216
1906	9592	2093	298	649	793	383
1916	15013	3184	403	1050	1369	339
1926	19065	3887	738	1273	1243	599
1936	25615	5063	>--2828--<		1441	641

verklaring: aanw.= aanwezig; opgen.= opgenomen
 (n) herst.= (niet) hersteld; overl.= overleden;
 accres = jaarlijkse toename.

Bij Koninklijk Besluit van 31 mei 1929 (Staatsblad nr. 275) worden de registers gewijzigd. Het belangrijkste verschil is het wegvallen van de vermelding bij ontslag: niet krankzinnig, niet hersteld of hersteld. In de jaarverslagen over 1930 en 1931 worden deze kwalificaties voor het laatst vermeld. In het 28e jaarverslag over de periode 1932-1936 (Pameyer, 1938) worden alleen het aantal opgenomen, ontslagen en overleden patiënten gepresenteerd. Dit jaarverslag is overigens het laatste dat in het Rijksarchief aanwezig is. Op 31 december 1936 bevinden zich 25.615 personen in inrichtingen, hetgeen 29,9 per 10.000 inwoners is. Het hoogste aantal tot nu toe. Het jaarlijkse accres is tot 641 gegroeid. In 1930 werd een recordtoename van 860 patiënten gemeld. Een dergelijke explosieve groei betekende dat de werkzaamheden van de inspecteurs zich dienden te wijzigen. Zag Schröder van der Kolk nog iedere patiënt en kon hij zich aldus heel direct op de hoogte stellen of de behandeling tot genezing strekte, in latere jaren was dit onmogelijk en beperkte de taak van de inspecteurs zich tot een strikt controlerende en toezichthoudende taak in meer algemene zin (Vijselaar, 1983). Het staatstoezicht op de krankzinnigenzorg richtte zich meer op de randvoorwaarden dan de directe zorg aan de patiënt. Nood-

gedwongen, immers in 1842 waren er twee part-time inspecteurs voor ca. 850 patiënten; in 1885 twee full-time inspecteurs voor 4.843 patiënten. In 1930 werd een derde full-time inspecteur benoemd, terwijl het aantal opgenomen patiënten inmiddels 4,5 maal groter was geworden en gestegen was tot 22.178, opgenomen in 42 inrichtingen.

Nauwgezet hebben de inspecteurs zich gekweten van hun taak *de bedoelingen van de wet te toetsen aan de werkelijke situatie*. Sinds het eerste verslag was in de jaarverslagen een hoofdstuk: "Werking van de Wet" opgenomen, waarin soms artikelsgewijze deze toetsing kwalitatief plaatsvond. Naast een beschrijving van alle veranderingen en eventuele misstanden binnen de diverse inrichtingen, is in de jaarverslagen steeds zeer grote aandacht besteed aan de cijfermatige gegevens van de opgenomen patiënten. Zo werd over bijna 100 jaar een uitstekende centrale registratie bijgehouden van alle patiënten met diverse persoonskenmerken. Zelfs evaluatieve gegevens werden verzameld: herstelquotiënt van de ontslagen patiënten, landelijk zowel als per inrichting. Het is te betreuren dat met deze gegevens zo weinig gedaan werd. Een handicap is ongetwijfeld geweest de omslachtige verwerking van deze gegevens met de hand en het gebrek aan statistische kennis voor een wetenschappelijkere beoordeling ervan. Vele onderzoekers zouden van een dergelijk gegevensbestand dromen: tal van vraagstellingen naar onderlinge relaties tussen de variabelen zouden onderzocht kunnen worden, alsmede epidemiologische vragen zouden ermee beantwoord kunnen worden. Het proefschrift van Kraus (1933) is erop gebaseerd.

De effectmeting: het herstelquotiënt is naar huidige maatstaven dubieus. Er valt niet te achterhalen waar de beoordeling *"wel of niet hersteld"* op is gebaseerd. Kennelijk vond men dat in die tijd een volstrekt evidente maat en gezien het type patiënten, enigszins begrijpelijk. Toch is dit een uniek evaluatief gegeven, dat, op de juiste wijze gebruikt, tot een landelijk evaluatie-instrument had kunnen worden uitgebreid. Geen enkele inspecteur heeft dit echter willen, durven of kunnen hanteren. Kwam dit doordat zij zelf wisten dat dit geen harde maatstaf was, en verdween het daarom in 1930 uit het register, of vreesden zij de directies van de inrichtingen die zo met elkaar vergeleken werden, of konden ze zelf met dit cijfermateriaal niet overweg en onthielden zij zich daarom van een inhoudelijke beschrijving ervan? Zelfs Schuurmans Stekhoven, die kennelijk een grote affiniteit had tot het verwerken van statistische gegevens (zie Schuurmans Stekhoven, 1922) geeft geen statistische c.q. toetsende beschrijving van dit materiaal. Was eenieder afgeschrikt door Schröder van der Kolk die in 1848 in het eerste jaarverslag schreef: *"... tevens (wordt) aangetoond, dat de statistiek, die wanneer zij met kennis van zaken beoordeeld en aangewend wordt, een krachtdadig en onwaardeerbaar middel is tot het aan het licht brengen van veel gewichtige waarheden, welke zonder haar verholten zouden blijven, bij eene niet oordeelkundige toepassing, tot een gevaarlijke gids kan worden*

gemaakt, die tot zeer valsche gevolgtrekkingen en tot grove dwalingen kan leiden" (p. 54). Het gemis aan 'oordeelkundige toepassing' zal zeker een rol gespeeld hebben.

In het jaar 1936 verschijnt het laatste jaarverslag van de inspecteurs. De opeenvolgende 28 verslagen waren de belangrijkste bron voor bovenstaande beschrijving. Dit is dan ook een kunstmatig einde van deze periode over het staatstoezicht op krankzinnigen. De programma-evaluatieve functie van dit toezicht is hiermee voldoende geadstrueerd.

1.7 VAN 1969 TOT EN MET 1984.

"Het stemt tot voldoening dat voor het eerst na bijna 40 jaar een verslag van de werkzaamheden van de Geneeskundige Inspectie voor de Geestelijke Volksgezondheid tot stand is gekomen", luidde de eerste zin van het jaarverslag dat in 1976 werd gepubliceerd en over de periode 1969-1975 handelde. De motieven voor dit intermezzo zijn slechts globaal aangeduid: de Tweede Wereldoorlog, grote maatschappelijke veranderingen, gewijzigde zorg aan psychiatrische patiënten, e.d. Dit eerste jaarverslag is geheel anders van opzet dan alle voorafgaande. Het is meer een beleidsmatige notitie over het functioneren van de totale geestelijke volksgezondheid, inclusief de ambulante zorg welke zich met name na 1940 sterk ontwikkelde zoals I.M.P., C.A.D., vele Sociaal Psychiatrische Diensten, e.d. De Geestelijke Volksgezondheid is steeds meer een lappendeken geworden van min of meer op elkaar aansluitende voorzieningen. In totaal bestaan er per 1-1-1974 194 extramurale voorzieningen.

De tabellen van patiëntengegevens, welke vóór 1936 een grote plaats innamen, beslaan nu slechts een klein gedeelte. De gepresenteerde gegevens zijn gebaseerd op het *Patiëntenregister Intramurale Geestelijke Gezondheidszorg (P.I.G.G.)*. In 1964 waren de voorbereidingen getroffen voor het opzetten van een landelijk registratiesysteem van alle opnamen in de psychiatrische ziekenhuizen. De bedoeling van dit systeem was het verzamelen van een kwantitatief inzicht in patiëntenstromen en patiëntenkarakteristieken (*Patiëntenregister Intramurale Geestelijke Gezondheidszorg*, 1985). Dit is een wezenlijk andere benadering dan het register van vóór 1936. Het is met behulp van P.I.G.G. mogelijk gegevens van een patiënt te volgen, waarbij ook heropnamen en overplaatsingen zichtbaar worden. Daarnaast kunnen nieuwe, computermatige verwerkingsmogelijkheden meer inzicht geven in onderlinge verbanden van diverse parameters. Per 1 januari 1967 werden de ontworpen registratiekaarten ingevoerd. Het belangrijkste verschil met het register anno 1884 is, dat de antwoorden bij de vragen geprecodeerd zijn, dit om automatische verwerking mogelijk te maken (zie afb. 1.7). In 1984 werd dit gewijzigd (zie afb. 1.8). Het belangrijkste verschil is dat

in het register van 1967 ook de verstrekte behandeling werd vastgelegd. Wellicht om reden dat deze informatie te vaak gemist zal worden is deze bij het register van 1984 verwijderd.

Per 1967 namen 39 psychiatrische ziekenhuizen aan deze registratie deel. Sindsdien zijn ook weer landelijke gegevens beschikbaar. Deze cijfers geven een grote verandering te zien welke zich de afgelopen 40 jaar, sinds 1936, heeft voltrokken. Waren in 1936 29,9 patiënten per 10.000 inwoners in de inrichtingen aanwezig, in 1967 waren dat er 20,6, een sterke relatieve daling. Absoluut gezien was het aantal hoegenaamd gelijk gebleven. (zie ook tabel 1.4)

De trend sinds het begin van de hier beschreven periode is (bron: P.I.G.G. jaarverslagen 1967 t/m 1983):

- een daling van het aantal aanwezige patiënten van 26.104 in 1967 naar ca. 21.700 patiënten in 1983 (van een algemeen psychiatrisch ziekenhuis ontbreken de gegevens). Relatief is de daling nog sterker: van 20,6 naar ca. 15 per 10.000 inwoners;
- een stijging van het aantal opnamen: van 11.596 naar ca. 25.000 ofwel van 9,2 naar ruim 17 per 10.000 inwoners;
- het aantal ontslagenen stijgt dienovereenkomstig;
- het aantal overleden patiënten blijft over deze periode constant: ca. 1500 per jaar;
- een sterke stijging van het aantal heropnamen. Sinds 1970 is het aantal heropnamen verdrievoudigd en beslaat in 1983 ca. 18% van alle opnemingen;
- de gemiddelde verblijfsduur daalde, althans van de nieuw opgenomen patiënten; in 1973 werd 69% binnen 6 maanden weer ontslagen, in 1983 was dit 79%;
- de bevolkingsopbouw in de psychiatrische ziekenhuizen veranderde eveneens: op 31 december 1967 was 19,8% minder dan 1 jaar opgenomen, in 1983 was dit 35,3%. Dit betekent ook dat er een grote groep verblijfspatiënten is ontstaan die reeds een of meer jaren zijn opgenomen en wiens kans op ontslag heel gering is, zeker naar een ambulante behandelingssituatie. Immers, op 31 december 1983 waren 65% van de patiënten meer dan een jaar opgenomen, waarvan wellicht slechts ca. 10% ontslagen zal worden in 1984.

Bovenstaande cijfermatige samenvatting besluit het kwantitatieve overzicht van 140 jaar formeel "*het staatstoezicht op krankzinnigen*". Maar ook kwalitatief zijn er aan het einde van deze lange periode enige opmerkelijke verschijnselen. Zoals hierboven beschreven, werden de psychiatrische ziekenhuizen voor ca. 55% bewoond door patiënten die hier langer dan 2 jaar verbleven, en volgens Van Weerden-Dijkstra (1976) tot het "chronische bestand" gerekend kunnen worden. In deze ziekenhuizen kunnen derhalve twee typen patiënten worden onderscheiden: zij die hier behandeld worden en binnen 1 jaar weer worden ontslagen - helaas keert hiervan weer ca. 18% terug - en de zgn. verblijfspatiënten. (zie ook Giel e.a., 1978; Haveman, 1980; Brook, 1985). In het jaarverslag van

34

[illegible]**VERVOLG OPNEMING**

<p>13 Is betrokkene opgenomen uit grote voorname?</p> <p><input type="checkbox"/> 1 ja <input type="checkbox"/> 2 neen</p> <p>45 <input type="checkbox"/></p> <p>Ze is uit</p> <p><input type="checkbox"/> 20 psych afw sig etc <input type="checkbox"/> 21 algemeen ziekthuis <input type="checkbox"/> 22 verpleeghuis <input type="checkbox"/> 23 verpleeghuis eigen kost <input type="checkbox"/> 24 psych ziekenhuis <input type="checkbox"/> 25 dagbesteding in psych etc <input type="checkbox"/> 26 overzorginstelling eigen inst <input type="checkbox"/> 27 overzorginstelling <input type="checkbox"/> 28 zelfde instelling <input type="checkbox"/> 29 t.b.v. -inrichting <input type="checkbox"/> 30 semiwettelijke voorziening <input type="checkbox"/> 31 kliniek voor verstandelijk en 32 creancetrum 33 gewone 34 huis van bewaring 35 anderszins nl</p> <p>56 <input type="checkbox"/></p> <p>Ze naam anders verbleeflocatie</p> <p><input type="checkbox"/> 10 afzondereinst <input type="checkbox"/> 11 pensioen/huis thuis <input type="checkbox"/> 12 ouderlijk huis <input type="checkbox"/> 13 eigen gezin <input type="checkbox"/> 14 familie/verrezen <input type="checkbox"/> 15 pleeggezin <input type="checkbox"/> 16 gastgezin <input type="checkbox"/> 17 anderszins nl</p>	<p>14 Is de opname (maakt) een gevolg van een (andere) gedrag?</p> <p><input type="checkbox"/> 1 ja <input type="checkbox"/> 2 neen</p> <p>54 <input type="checkbox"/></p> <p>Ze is uit de opname (maakt) een gevolg van een (andere) gedrag?</p> <p><input type="checkbox"/> 1 ja <input type="checkbox"/> 2 neen</p> <p>55 <input type="checkbox"/></p> <p>Ze is uit de opname (maakt) een gevolg van een (andere) gedrag?</p> <p><input type="checkbox"/> 1 ja <input type="checkbox"/> 2 neen</p> <p>56 <input type="checkbox"/></p> <p>Ze is uit de opname (maakt) een gevolg van een (andere) gedrag?</p> <p><input type="checkbox"/> 1 ja <input type="checkbox"/> 2 neen</p> <p>57 <input type="checkbox"/></p> <p>Ze is uit de opname (maakt) een gevolg van een (andere) gedrag?</p> <p><input type="checkbox"/> 1 ja <input type="checkbox"/> 2 neen</p> <p>58 <input type="checkbox"/></p> <p>Ze is uit de opname (maakt) een gevolg van een (andere) gedrag?</p> <p><input type="checkbox"/> 1 ja <input type="checkbox"/> 2 neen</p> <p>59 <input type="checkbox"/></p> <p>Ze is uit de opname (maakt) een gevolg van een (andere) gedrag?</p> <p><input type="checkbox"/> 1 ja <input type="checkbox"/> 2 neen</p> <p>60 <input type="checkbox"/></p> <p>Ze is uit de opname (maakt) een gevolg van een (andere) gedrag?</p> <p><input type="checkbox"/> 1 ja <input type="checkbox"/> 2 neen</p> <p>61 <input type="checkbox"/></p> <p>Ze is uit de opname (maakt) een gevolg van een (andere) gedrag?</p> <p><input type="checkbox"/> 1 ja <input type="checkbox"/> 2 neen</p> <p>62 <input type="checkbox"/></p> <p>Ze is uit de opname (maakt) een gevolg van een (andere) gedrag?</p> <p><input type="checkbox"/> 1 ja <input type="checkbox"/> 2 neen</p> <p>63 <input type="checkbox"/></p> <p>Ze is uit de opname (maakt) een gevolg van een (andere) gedrag?</p> <p><input type="checkbox"/> 1 ja <input type="checkbox"/> 2 neen</p> <p>64 <input type="checkbox"/></p> <p>Ze is uit de opname (maakt) een gevolg van een (andere) gedrag?</p> <p><input type="checkbox"/> 1 ja <input type="checkbox"/> 2 neen</p> <p>65 <input type="checkbox"/></p> <p>Ze is uit de opname (maakt) een gevolg van een (andere) gedrag?</p> <p><input type="checkbox"/> 1 ja <input type="checkbox"/> 2 neen</p> <p>66 <input type="checkbox"/></p> <p>Ze is uit de opname (maakt) een gevolg van een (andere) gedrag?</p> <p><input type="checkbox"/> 1 ja <input type="checkbox"/> 2 neen</p> <p>67 <input type="checkbox"/></p> <p>Ze is uit de opname (maakt) een gevolg van een (andere) gedrag?</p> <p><input type="checkbox"/> 1 ja <input type="checkbox"/> 2 neen</p> <p>68 <input type="checkbox"/></p> <p>Ze is uit de opname (maakt) een gevolg van een (andere) gedrag?</p> <p><input type="checkbox"/> 1 ja <input type="checkbox"/> 2 neen</p> <p>69 <input type="checkbox"/></p> <p>Ze is uit de opname (maakt) een gevolg van een (andere) gedrag?</p> <p><input type="checkbox"/> 1 ja <input type="checkbox"/> 2 neen</p> <p>70 <input type="checkbox"/></p> <p>Ze is uit de opname (maakt) een gevolg van een (andere) gedrag?</p> <p><input type="checkbox"/> 1 ja <input type="checkbox"/> 2 neen</p> <p>71 <input type="checkbox"/></p> <p>Ze is uit de opname (maakt) een gevolg van een (andere) gedrag?</p> <p><input type="checkbox"/> 1 ja <input type="checkbox"/> 2 neen</p> <p>72 <input type="checkbox"/></p> <p>Ze is uit de opname (maakt) een gevolg van een (andere) gedrag?</p> <p><input type="checkbox"/> 1 ja <input type="checkbox"/> 2 neen</p> <p>73 <input type="checkbox"/></p> <p>Ze is uit de opname (maakt) een gevolg van een (andere) gedrag?</p> <p><input type="checkbox"/> 1 ja <input type="checkbox"/> 2 neen</p> <p>74 <input type="checkbox"/></p> <p>Ze is uit de opname (maakt) een gevolg van een (andere) gedrag?</p> <p><input type="checkbox"/> 1 ja <input type="checkbox"/> 2 neen</p> <p>75 <input type="checkbox"/></p> <p>Ze is uit de opname (maakt) een gevolg van een (andere) gedrag?</p> <p><input type="checkbox"/> 1 ja <input type="checkbox"/> 2 neen</p> <p>76 <input type="checkbox"/></p> <p>Ze is uit de opname (maakt) een gevolg van een (andere) gedrag?</p> <p><input type="checkbox"/> 1 ja <input type="checkbox"/> 2 neen</p> <p>77 <input type="checkbox"/></p> <p>Ze is uit de opname (maakt) een gevolg van een (andere) gedrag?</p> <p><input type="checkbox"/> 1 ja <input type="checkbox"/> 2 neen</p> <p>78 <input type="checkbox"/></p> <p>Ze is uit de opname (maakt) een gevolg van een (andere) gedrag?</p> <p><input type="checkbox"/> 1 ja <input type="checkbox"/> 2 neen</p> <p>79 <input type="checkbox"/></p> <p>Ze is uit de opname (maakt) een gevolg van een (andere) gedrag?</p> <p><input type="checkbox"/> 1 ja <input type="checkbox"/> 2 neen</p> <p>80 <input type="checkbox"/></p> <p>Ze is uit de opname (maakt) een gevolg van een (andere) gedrag?</p> <p><input type="checkbox"/> 1 ja <input type="checkbox"/> 2 neen</p> <p>81 <input type="checkbox"/></p> <p>Ze is uit de opname (maakt) een gevolg van een (andere) gedrag?</p> <p><input type="checkbox"/> 1 ja <input type="checkbox"/> 2 neen</p> <p>82 <input type="checkbox"/></p> <p>Ze is uit de opname (maakt) een gevolg van een (andere) gedrag?</p> <p><input type="checkbox"/> 1 ja <input type="checkbox"/> 2 neen</p> <p>83 <input type="checkbox"/></p> <p>Ze is uit de opname (maakt) een gevolg van een (andere) gedrag?</p> <p><input type="checkbox"/> 1 ja <input type="checkbox"/> 2 neen</p> <p>84 <input type="checkbox"/></p> <p>Ze is uit de opname (maakt) een gevolg van een (andere) gedrag?</p> <p><input type="checkbox"/> 1 ja <input type="checkbox"/> 2 neen</p> <p>85 <input type="checkbox"/></p> <p>Ze is uit de opname (maakt) een gevolg van een (andere) gedrag?</p> <p><input type="checkbox"/> 1 ja <input type="checkbox"/> 2 neen</p> <p>86 <input type="checkbox"/></p> <p>Ze is uit de opname (maakt) een gevolg van een (andere) gedrag?</p> <p><input type="checkbox"/> 1 ja <input type="checkbox"/> 2 neen</p> <p>87 <input type="checkbox"/></p> <p>Ze is uit de opname (maakt) een gevolg van een (andere) gedrag?</p> <p><input type="checkbox"/> 1 ja <input type="checkbox"/> 2 neen</p> <p>88 <input type="checkbox"/></p> <p>Ze is uit de opname (maakt) een gevolg van een (andere) gedrag?</p> <p><input type="checkbox"/> 1 ja <input type="checkbox"/> 2 neen</p> <p>89 <input type="checkbox"/></p> <p>Ze is uit de opname (maakt) een gevolg van een (andere) gedrag?</p> <p><input type="checkbox"/> 1 ja <input type="checkbox"/> 2 neen</p> <p>90 <input type="checkbox"/></p> <p>Ze is uit de opname (maakt) een gevolg van een (andere) gedrag?</p> <p><input type="checkbox"/> 1 ja <input type="checkbox"/> 2 neen</p> <p>91 <input type="checkbox"/></p> <p>Ze is uit de opname (maakt) een gevolg van een (andere) gedrag?</p> <p><input type="checkbox"/> 1 ja <input type="checkbox"/> 2 neen</p> <p>92 <input type="checkbox"/></p> <p>Ze is uit de opname (maakt) een gevolg van een (andere) gedrag?</p> <p><input type="checkbox"/> 1 ja <input type="checkbox"/> 2 neen</p> <p>93 <input type="checkbox"/></p> <p>Ze is uit de opname (maakt) een gevolg van een (andere) gedrag?</p> <p><input type="checkbox"/> 1 ja <input type="checkbox"/> 2 neen</p> <p>94 <input type="checkbox"/></p> <p>Ze is uit de opname (maakt) een gevolg van een (andere) gedrag?</p> <p><input type="checkbox"/> 1 ja <input type="checkbox"/> 2 neen</p> <p>95 <input type="checkbox"/></p> <p>Ze is uit de opname (maakt) een gevolg van een (andere) gedrag?</p> <p><input type="checkbox"/> 1 ja <input type="checkbox"/> 2 neen</p> <p>96 <input type="checkbox"/></p> <p>Ze is uit de opname (maakt) een gevolg van een (andere) gedrag?</p> <p><input type="checkbox"/> 1 ja <input type="checkbox"/> 2 neen</p> <p>97 <input type="checkbox"/></p> <p>Ze is uit de opname (maakt) een gevolg van een (andere) gedrag?</p> <p><input type="checkbox"/> 1 ja <input type="checkbox"/> 2 neen</p> <p>98 <input type="checkbox"/></p> <p>Ze is uit de opname (maakt) een gevolg van een (andere) gedrag?</p> <p><input type="checkbox"/> 1 ja <input type="checkbox"/> 2 neen</p> <p>99 <input type="checkbox"/></p> <p>Ze is uit de opname (maakt) een gevolg van een (andere) gedrag?</p> <p><input type="checkbox"/> 1 ja <input type="checkbox"/> 2 neen</p> <p>100 <input type="checkbox"/></p> <p>Ze is uit de opname (maakt) een gevolg van een (andere) gedrag?</p> <p><input type="checkbox"/> 1 ja <input type="checkbox"/> 2 neen</p>	<p>15 Is de opname (maakt) een gevolg van een (andere) gedrag?</p> <p><input type="checkbox"/> 1 ja <input type="checkbox"/> 2 neen</p> <p>54 <input type="checkbox"/></p> <p>Ze is uit de opname (maakt) een gevolg van een (andere) gedrag?</p> <p><input type="checkbox"/> 1 ja <input type="checkbox"/> 2 neen</p> <p>55 <input type="checkbox"/></p> <p>Ze is uit de opname (maakt) een gevolg van een (andere) gedrag?</p> <p><input type="checkbox"/> 1 ja <input type="checkbox"/> 2 neen</p> <p>56 <input type="checkbox"/></p> <p>Ze is uit de opname (maakt) een gevolg van een (andere) gedrag?</p> <p><input type="checkbox"/> 1 ja <input type="checkbox"/> 2 neen</p> <p>57 <input type="checkbox"/></p> <p>Ze is uit de opname (maakt) een gevolg van een (andere) gedrag?</p> <p><input type="checkbox"/> 1 ja <input type="checkbox"/> 2 neen</p> <p>58 <input type="checkbox"/></p> <p>Ze is uit de opname (maakt) een gevolg van een (andere) gedrag?</p> <p><input type="checkbox"/> 1 ja <input type="checkbox"/> 2 neen</p> <p>59 <input type="checkbox"/></p> <p>Ze is uit de opname (maakt) een gevolg van een (andere) gedrag?</p> <p><input type="checkbox"/> 1 ja <input type="checkbox"/> 2 neen</p> <p>60 <input type="checkbox"/></p> <p>Ze is uit de opname (maakt) een gevolg van een (andere) gedrag?</p> <p><input type="checkbox"/> 1 ja <input type="checkbox"/> 2 ne</p>
---	---	--

Afb. 1.7 P.I.G.G. registratiekaart van 1984
(Overgenomen met toestemming van de hoofdinspectie)

Tabel 1.4 Overzicht van opnamen en ontslagen over de periode 1967 - 1983 in intervallen van 4 jaar

jaar	aanw.	opg.	ontsl.	overl.	accres
1967	26104	11596	9927	1570	252
1971	25685	13671	12279	1700	- 213
1975	23695	18534	17425	1482	- 339
1979a)	21821	19280	19674	1381	- 30
1983b)	21406	24781	23343	1592	-

a) excl. W.A.Huis te Utrecht en Zon en Schild te Amersfoort, ca. 1100 pat.

b) excl. W.A.Huis te Utrecht, ca. 325 patiënten.

de Geneeskundige (Hoofd-)Inspectie voor de Geestelijke Volksgezondheid over 1983 wordt de kritiek van de hand gewezen, dat dit te wijten zou zijn aan een slecht behandelingsresultaat (p. 96). De inspectie is van mening, dat onderzocht moet worden of deze verblijfspatiënten niet in een apart centrum ondergebracht moeten worden. In feite is hier dan weer sprake van een *"bewaarplaats nieuwe stijl"*, waar anno 1850 reeds discussie over was (zie pag. 27). Door het scheiden van behandel- en verblijfspatiënten zullen dan ook de statistieken van de Algemeen Psychiatrische Ziekenhuizen een ander beeld geven. En dat idee is ook reeds in 1822 geopperd, toen men discussiëerde over de scheiding van de *"geneeslijken en ongeneeslijken"*, dit in het kader van de *"eigenliefde"* van de arts (zie p. 13). Zowel aan het verschijnsel *"ongeneeslijken"* als aan dat van de heropnamen wordt verder geen aandacht besteed, terwijl dit toch juist het belangrijkste argument lijkt om de psychiatrische ziekenhuizen een gering behandelingsresultaat te verwijten. Voor dit intramurale deel van de geestelijke gezondheidszorg zou een 'motie Beckers' en een hierop aansluitend onderzoek (Schagen, 1985) eveneens op zijn plaats zijn.

Een andere belangrijke ontwikkeling is het bezuinigingsbeleid dat sedert 1977 in toenemende mate ook in de geestelijke gezondheidszorg wordt uitgevoerd. Het jaarverslag over 1978 meldde nog dat kwaliteit van zorg samengaat met de beschikbare middelen en dat deze zorg prioriteit heeft bij het gevoerde ombuigingsbeleid. Uit de jaarverslagen van 1982 en 1983 blijkt van het ontzien van deze geestelijke gezondheidszorg weinig: een vacaturestop, 2% ambtenarenreductie, administratieve bouwstop en geen personeelsuitbreiding voor intensieve behandelingsvormen, e.d. zijn bezuinigende maatregelen die het tegendeel aangeven. Het laatste jaar van dit overzicht van de staatsbemoeienis, gestart in feite in 1816 bij Willem I die de erbarmelijke toestand van de krankzinnigen wilde verbeteren, is dan 1984. Het verslag over dit jaar leidde tot een krantenkop: *"Psychiatrische patiënt door*

geldgebrek verwaarloosd" (Volkskrant, 2-11-1985). Dit jaarverslag meldde dat: "... in een toenemend aantal klinische afdelingen van APZ'en weer verwaarlozingsverschijnselen bij de patiënten kunnen worden waargenomen door een te geringe verpleegkundige bezetting" (p. 92). Het is wrang dat na 168 jaar de cirkel zich lijkt te sluiten. Uiteraard is de term "verwaarlozing" anno 1984 van een andere kwaliteit dan die van 1816. Maar een passage "patiënt bevindt zich wekenlang ongekleed in een isoleercel" zou ook aangetroffen kunnen worden in de eerste jaarverslagen, alsmede uitspraken zoals "onaanvaardbare slechte huisvestingssituatie" (p. 40). Zoals aan het begin van dit historisch overzicht reeds werd afgevraagd, moet worden geconstateerd dat de geschiedenis zich lijkt te herhalen, en er is dus weer te weinig van geleerd.

Een andere constatering is dat de programma-evaluatieve taak van de huidige inspecteurs veel afstandelijker en in meer algemenere zin werd uitgevoerd dan door hun collega's uit de vorige eeuw. Dit is jammer, permanent toezicht en feedback heeft in het verleden de zorg voor de 'krankzinnige' zeer verbeterd, ook in periodes dat 's lands schatkist eveneens weinig gevuld was. Zou het ontbreken van deze aanhoudende evaluatie en de verslechterende situatie van de psychiatrische patiënt niet met elkaar in verband staan, en/of had een dergelijke evaluatie dit kunnen voorkomen?

1.8 VAN VERLEDEN NAAR TOEKOMST.

Het staatstoezicht op de krankzinnigenzorg is genoemd: *"omzien naar weldoen"*. In bijna 170 jaar is veel gedaan om de zorg voor de krankzinnige c.q. de psychiatrische patiënt te verbeteren, de behandeling, de huisvesting, de dagelijkse verzorging, etc. Deze verbeteringen zijn mede bewerkstelligd door de voortdurende en nauwgezette controle van de inspecteurs; als eerste programma-evaluatoren in de geestelijke gezondheidszorg hebben zij de zorg voor de psychiatrische patiënt herhaaldelijk getoetst aan de doelstellingen. De evaluatieve maat: (niet) verbeterd, hersteld ontslagen heeft hierbij een (te) bescheiden rol gespeeld; maar deze was ook niet nodig in een tijd dat de inspecteurs zich persoonlijk van het functioneren van de behandeling en de effectiviteit ervan op de hoogte konden stellen. Het is des te opmerkelijker dat deze maat werd afgeschaft toen deze juist hard nodig was: de resultaten werden minder en directe controle en toezicht was niet meer mogelijk door het grote aantal patiënten. Een dergelijke evaluatieve maat zou weer ontwikkeld en ingevoerd moeten worden. Een tweede conclusie is, dat directe feedback op instellingsniveau een belangrijk sturend mechanisme is geweest. Het slechts landelijk presenteren van registratiegegevens, zoals dit de laatste jaren gebeurt, heeft deze invloed niet. Ten derde hadden de inspecteurs geen machtsmiddelen om bij geconstateerde gebreken op te tre-

den. Hun taak was rapportierend. Desalniettemin zijn (mede) door hun informatie de noodzakelijke veranderingen aangebracht, wellicht in een trager tempo dan zij dit zelf wenselijk hebben geacht. Hun informatie werd getoetst aan een doel: de wet van 1841. En deze diende toch te worden nageleefd. Daarnaast hebben zij actief bijgedragen aan het verbeteren van deze doelstellingen door hun voorstellen voor de wet van 1884. Hun descriptieve taak werd zo ruimer dan aanvankelijk bedoeld. Ze hadden nu ook een adviserende functie, en gelet op hun grote kennis van zaken, terecht.

Hiermee zijn de belangrijkste elementen voor een (landelijke) programma-evaluatie gegeven:

- toetsbare doelstellingen, welke algemeen geaccepteerd zijn;
- een hanteerbare evaluatieve maat;
- feedback van de resultaten op extern en intern niveau;
- waarborgen dat de uitkomsten van het evaluatieproces ook ertoe zullen leiden dat de behandeling verbetert;
- een zodanige opzet en rolverdeling van de participanten dat een bijdrage geleverd wordt aan het totale evaluatieproces zelf, waardoor de mankementen aan doelen, metingen en feedback worden verbeterd.

In evaluatief opzicht is dit de lering die uit de geschiedenis van het staatstoezicht op de krankzinnigenzorg kan worden getrokken. De invloed die programma-evaluatie zou kunnen leveren, wordt in toenemende mate, ook op ministeriële niveau, erkend middels een stimuleringsbeleid voor dit type onderzoek (zie vergaderstuk 18161, 2e Kamer der Staten Generaal, en Dekker, 1984).

Hopelijk kan het werk dat door koning Willem I is begonnen, ook nu nog worden gecontinueerd, met aan deze tijd aangepaste middelen. *Programma-evaluatie kan zo'n middel zijn.*

Hoofdstuk 2:

Is evalueren ook onderzoeken?

*Onderzoek alles en behoud het goede,
Evalueer het om het te verbeteren.*

2.1 INLEIDING.

In het boek *"Evaluation Research and Practice"* van Levine e.a. (1981) begint het eerste hoofdstuk aldus: *"In terms of world culture, perhaps the two most distinctive U.S. contributions of the 1970's are the movie Star Wars and evaluation research. There are many similarities between them: Both are a mixture of reality and fantasy, both have proved lucrative, both have been critically examined and acclaimed by other nations, and both will be refined and explained in concept during the next decade"*. Deze enigszins bombastische beschrijving wordt relativerend afgesloten met de opmerking dat beide bijdragen geen nieuwe uitvindingen zijn: in 1930 waren er in de U.S. reeds science fiction films vergelijkbaar met "Star Wars" en volgens Levine voert evaluatie-onderzoek terug naar de sociale experimenten van Kurt Lewin. Ten aanzien van "Star Wars" kan de oorsprong wellicht nog terecht zijn, de relatie tussen evaluatie-onderzoek en Kurt Lewin lijkt minder juist. Lewin onderzocht sociaal psychologische fenomenen middels sociale experimenten en ontwikkelde o.a. zijn "veld-theorie" (Deutsch en Krauss, 1965; Lewin, 1967). Dit had weinig van doen met evaluatie-onderzoek, althans niet in de context van genoemd boek. Ten aanzien van de verdere ontwikkeling van evaluatie-research mag overigens gehoopt worden dat deze niet dezelfde weg zal gaan als "Star Wars" c.q. S.D.I.

De oorsprong van evaluatie-research wordt door andere auteurs veel eerder gelegd: in de vorige eeuw (Schnabel, 1984a) of zelfs in het *Oude Testament* (Patton, 1980; Blöte e.a., 1983). Dit oudste evaluatie-onderzoek is het vermelden waard. Het moet geplaatst worden ca. 600 v. Chr. Koning Nebukadnezar, koning van Babel, veroverde Jeruzalem en nam vier Israëliëten van koninklijke bloede mee naar Babel en liet ze gedurende drie jaar aan het Babylonische hof opvoeden, waar ze de taal der Chaldeeën leerden om vervolgens in dienst van de koning te komen. Onder hen bevonden zich enige Judeeërs, met name Daniël. Zij dienden ook van de koninklijke dis te eten. De bijbeltekst luidt verder als volgt:

Daniël nu nam zich voor zich niet te verontreinigen met de koninklijke spijsen of met de wijn die de koning placht te drinken en hij verzocht de overste der hovelingen dat hij zich niet zou behoeven te verontreinigen. Toen schonk God aan Daniël gunst en barmhartigheid bij de overste der hovelingen; doch de overste der hovelingen zei tot Daniël: Ik vrees dat mijn heer, de koning, die uw spijs en drank vastgesteld heeft, uw uiterlijk minder welvarend zou vinden dan dat van de overige knapen van uw leeftijd, en dat gij daardoor mijn hoofd met schuld zoudt beladen bij de koning. Daarop zei Daniël (.....): Neem toch met uw dienaren gedurende tien dagen de proef: men geve ons groenten te eten en water te drinken; laat dan ons uiterlijk met dat van de knapen die de koninklijke spijsen eten, door u vergeleken worden, en doe dan met uw dienaren naar uw bevindingen. Hij gaf hen hierin gehoor en nam met hen gedurende tien dagen de proef, en na verloop van tien dagen bleek hun uiterlijk schoner en zagen zij er welvarender uit dan al de knapen die van de koninklijke spijs gegeten hadden. Voortaan nam de kamerdienaar hun spijs en wijn die ze moesten eten en drinken weg en gaf hen groenten. En aan deze vier gaf God kennis en verstand van allerlei geschriften en wijsheid, terwijl Daniël inzicht had in allerlei gezichten en dromen.(.....) In elke zaak, waarbij het aankwam op wijs inzicht en waarover de koning hen ondervroeg, bevond hij hen tienmaal voortreffelijker dan al de geleerden, al de bezweerders in zijn ganse rijk"

Het oude testament, boek Daniël 1: 8-21.

Inderdaad is hier sprake van een *vergelijkend (evaluatief) onderzoek*: er is een *experimentele* (vegetarisch eten) groep en een *controlegroep*. De vraagstelling is of vegetarisch eten even goede effecten heeft als het eten van de koninklijke dis. De kamerdienaar was de beoordelaar van het effect: het uiterlijk van Daniël en de zijnen. Vooraf was afgesproken wat er diende te gebeuren met de uitslag. Het resultaat viel gunstig uit voor de experimentele groep: deze zag er beter uit dan de controlegroep. Daarnaast waren er voor Daniël c.s. nog enige positieve neveneffecten op de lange termijn (kennis, e.d.), welke eveneens gerapporteerd werden. Deze eerste wetenschappelijke onderzoeker belandde uiteindelijk ook nog eens in de leeuwenkuil en hem werd het vuur na aan de schenen gelegd, maar wist ook deze penibele situaties te overwinnen.

Dit onderzoek toont tevens aan, dat de gevonden effecten niet alle oorzakelijk aan de experimentele factor toegeschreven kunnen worden (niet méér kennis door het eten van groente), maar dat er van een onderliggende oorzaak sprake is

(gehoorzaamheid aan God die hen daarom die kennis gaf). Blöte (1983) citeert ten onrechte, dat "voortaan alle kinderen vegetarisch voedsel kregen". Dit is niet in overeenstemming met de oorspronkelijke tekst, maar het zou ook een onjuiste conclusie zijn, omdat niet het vegetarische voedsel de bepalende factor was. Indien wel, dan zou dit het eetpatroon tot op heden drastisch hebben veranderd. Het onderzoek werd wel benut en had directe consequenties voor de betrokkenen: Daniël en zijn makkers, zonder dat de door Blöte vermelde oneigenlijke generalisaties plaatsvonden.

Of hier nu sprake is van een evaluatie-onderzoek is een definitiekwestie: wat moet worden verstaan onder "*onderzoek*", "*evaluatie*", "*evaluatie-onderzoek*" en later te vermelden "*programma-evaluatie*"? Dit definitieprobleem, en met name of programma-evaluatie nu wel of geen (wetenschappelijk) onderzoek moet worden genoemd, wordt in academische kring reeds vele jaren gevoerd. De meeste boeken over programma-evaluatie wijden hier aandacht aan (Suchman, 1967, p. 31; Anderson en Ball, 1978, p. 9; Rutman, 1977, p. 45 e.v., Rutman, 1980, p. 18; Levine e.a., 1981, p. 24).

In dit hoofdstuk zal worden aangetoond, dat programma-evaluatie zelfs tot de *kern* van het wetenschappelijk onderzoek kan worden gerekend. Hiertoe wordt eerst in de volgende paragraaf het begrip onderzoek en zijn verschijningsvormen nader omschreven, teneinde in paragraaf 2.3 te komen tot een onderscheid tussen onderzoek en evaluatie. In paragraaf 2.4 worden tenslotte verschillende vormen van evaluatie beschreven, uitmondend in het type evaluatie dat in dit boek centraal staat: *programma-evaluatie*.

2.2 WAT IS ONDERZOEK?

Volgens Van Dale (1984) is *onderzoek*: "... de behandeling van het onderzoeken teneinde door de verzameling van gegevens tot een oplossing van een bepaald probleem te komen". Deze definitie is dusdanig breed, dat iedere activiteit waarin op een (impliciet gestelde) vraag een antwoord getracht wordt te geven, als onderzoek kan worden bestempeld. Wensen wij een meer wetenschappelijke definitie van onderzoek, dan lijkt een logische stap het raadplegen van De Groot's "*Methodologie*" (1961), voor vele gedragswetenschappers die eind jaren zestig zijn opgeleid de "*methodologie-bijbel*". De Groot gaat uit van een grondvorm voor onderzoek: *het toetsingsonderzoek*, "... d.i. het type onderzoek, waarbij een of enkele uit een theorie afgeleide hypothesen aan de empirie worden getoetst" (p. 315). Naast dit model beschrijft hij nog vier andere vormen van onderzoek. De hoofdkenmerken van deze vijf typen worden hier kort weergegeven.

1. *Toetsingsonderzoek*: een beperkt aantal gewoonlijk aan een theorie ontleende, onderling samenhangende hypothesen worden aan empirisch materiaal getoetst. Meestal betreft het

een steekproefonderzoek, dat vaak, zij het lang niet altijd, onder experimentele condities wordt uitgevoerd. Als voorbeelden van dergelijk onderzoek noemt De Groot: het toetsen van een bestaande theorie aan een of meer van haar consequenties. Als een theorie niet voorhanden is, kan men bijv. het bestaan van een effect proberen aan te tonen als uitgangspunt voor verdere theorievorming en empirisch onderzoek. Hij verwijst hierbij naar het zgn. *evaluatieprobleem*: "*de objectieve evaluatie van de effecten van methoden die erop gericht zijn het menselijk gedrag te beïnvloeden*" (p. 182). In toetsingsonderzoek dienen vooraf (falsifiëerbare) hypothesen gesteld te worden hetgeen "... dwingt tot explicitering, tot objectiviteit, tot vermindering van contaminatie, tot het nemen van risico's" (p. 318).

2. *Instrumenteel-nomologisch onderzoek* heeft tot doel het ontwikkelen van een instrument. De kenmerken zijn dat gewerkt wordt binnen en aan het nomologisch netwerk rond een bepaald begrip. Essentieel is dat getracht wordt dit begrip meetbaar te maken middels een instrument. Te denken valt aan de constructie van een vragenlijst voor het meten van bijv. depressie of neuroticisme.
3. *Descriptief onderzoek* heeft, vergeleken met de voorafgaande twee typen slechts een beperkte doelstelling: het beschrijven volgens een bepaalde systematiek van datgene wat zich op een bepaald terrein voordoet. Dit type is vaak een voorfase voor toetsend onderzoek.
4. *Exploratief onderzoek* is bedoeld om hypothesen te genereren. Chronologisch kan het een plaats vinden tussen descriptief en toetsend onderzoek.
5. *Interpretatief theoretische studies*. Dit type is een niet-empirische vorm van onderzoek, er worden geen nieuwe gegevens verzameld. Op een reeds bestaande, afgeronde of gesloten gegevensverzameling worden nadere (ad hoc) gestelde hypothesen getoetst c.q. verbanden nader geïnterpreteerd. Dit kan echter uitermate subjectief zijn; het resultaat, de interpretatie, is achteraf op validiteit niet te controleren.

Uit deze opsomming blijkt dat het toetsende onderzoek als het *basismodel* voor onderzoek wordt beschouwd. Het instrumenteel-nomologische onderzoek kan ten dienste staan van dit toetsend onderzoek: het verschaft (gekwaliceerde) operationalisaties. Descriptief en exploratief onderzoek zijn voorstadia: in descriptief onderzoek wordt het onderzoeksterrein "verkend", in exploratief onderzoek wordt dit vervolgens "bouwrijp" gemaakt. Het vijfde type staat er - ook volgens De Groot - enigszins buiten. Hij noemt dit dan ook een "niet empirisch" onderzoek, omdat er geen nieuwe gegevens verzameld worden.

Uit dit overzicht van onderzoeksvormen blijkt tevens, dat De Groot het evalueren van effecten van een behandeling rekent tot het basismodel van onderzoek. Ten onrechte wordt dan ook door sociale wetenschappers onderzoek dat gebruik

maakt van experimentele condities als de enige vorm van (wetenschappelijk) onderzoek bestempeld en al het andere als descriptief of "zacht" onderzoek gediskwalificeerd. De discussie of (programma-)evaluatie al dan niet onderzoek is, had niet gevoerd hoeven worden wanneer De Groot niet zo beperkt was geïnterpreteerd: evalueren is zeker onderzoeken. En het is *wetenschappelijk* als het dwingt tot expliciteren en tot objectiviteit en daarentegen contaminaties vermijdt.

Een andere corripce op het gebied van de methodologie is Hofstee. In zijn boek: *De empirische discussie* (1980) geeft hij een duidelijke opvatting over welke onderzoeksmodellen binnen de "empirische discussie" een plaats hebben. Hij gebruikt het "*weddenschapsmodel*" als grondvorm voor wetenschappelijk onderzoek. Onderzoek is dan bedoeld om aan te geven welke van de (minstens) twee partijen *gelijk* heeft. Deze partijen hebben dus een meningsverschil (eten van de koninklijke dis is wel/niet goed, er moeten wel/niet mensen in een bedrijf ontslagen worden, de hulpverlening heeft wel/geen succes). Partijen kunnen identificeerbare personen zijn (Daniël versus de overste der hovelingen), verzamelingen van personen of instanties (directie versus ondernemingsraad c.q. vakbonden) of conflicterende algemeen geldende meningen of opvattingen. In dit laatste geval adviseert Hofstee (1980, p. 115) de partijen te concretiseren. Zo kan ook de onderzoeker zelf de representant van een van de partijen zijn, en anderen die door uitspraken hiervan blijf hebben gegeven, de publieke opinie vertegenwoordigen. Of de onderzoeker vertegenwoordigt de publieke opinie versus degenen wier functioneren ter discussie staat, een situatie die zich bij externe programma-evaluatie kan voordoen (hulpverleners versus publieke opinie i.c. de evaluator).

De belangrijkste onderzoeksvariant van het weddenschapsmodel is het (*kritische*) *experiment* met als bijzonder geval het *evaluatie-onderzoek*. Over dit laatste zegt Hofstee: "*Voor zover iemand met betrekking tot een nieuw product of procedé gegevens verzamelt, wordt dat wel als een deelname aan de empirische discussie geïnterpreteerd: zulk evaluatie-onderzoek waarin geprobeerd wordt (potentiële) tegensprekers te overtuigen van de waarde van het maaksel staat in het model zelfs centraal*" (p. 119). Hiermee staan Hofstee en De Groot op een lijn: het evaluatie-onderzoek komt overeen met het zgn. evaluatieprobleem (De Groot, 1961, p. 182). De onderzoekstypen welke De Groot beschrijft, zoals: instrumenteel-nomologisch, descriptief en exploratief onderzoek, worden door Hofstee als "investeringsonderzoek" bestempeld en aan de rand van het weddenschapsmodel geplaatst. In feite bestaat er in deze geen verschil tussen De Groot en Hofstee: beiden hanteren het empirisch-toetsende onderzoek als model voor het wetenschappelijke onderzoeksbedrijf. Evaluatie-onderzoek is hiervan een speciale vorm die door beiden in het centrum van het model wordt geplaatst.

Hofstee onderscheidt verder nog evaluatie-onderzoek van diagnostisch onderzoek. Dit onderscheid wordt ook wel aangegeven met de termen: "toegepast versus theoretisch" of

"decisie versus conclusiegericht" onderzoek. Bij diagnostisch onderzoek wordt op basis van de onderzoeksresultaten een beslissing genomen, er wordt gekozen tussen twee alternatieven, bijv. wel of niet antidepressiva voorschrijven. Dit onderzoek komt in zijn zuivere vorm nauwelijks voor, omdat de gebruikte meetinstrumenten doorgaans geen absoluut criterium hebben op basis waarvan een beslissing kan worden genomen. Tal van andere, zoals emotionele argumenten zullen eveneens een rol spelen, en meestal ook terecht. Bij evaluatieonderzoek is er reeds een beslissing genomen. Nagegaan moet worden of deze beslissing juist was, of - in het weddenschapmodel geformuleerd - welke van de twee partijen: de voor- en tegenstanders van de beslissing, gelijk krijgen. Het belang van dergelijk onderzoek wordt door Hofstee vooral gezien in een theoretische zuiver wetenschappelijke situatie eerder dan in de toepassingspraktijk. Volgens hem worden er toch geen conclusies aan de resultaten verbonden, er kan hoogstens een preventieve werking van uitgaan wanneer de partijen zich iets van hun ongelijk aantrekken. Te vaak wordt het doen van evaluatie-onderzoek alleen maar gebruikt om eigen ideeën te bevestigen (verkiezingsonderzoek), om tijd te winnen of om indruk op anderen te maken (public relations).

Dit oneigenlijke gebruik van evaluatie bedreigt de empirische discussie, aldus Hofstee (p. 123), omdat het niet gebeurt om ongelijk te bekennen c.q. gelijk te krijgen. Sterker nog, volgens Hofstee wil men dikwijls niet eens kennis nemen van de resultaten. De onderzoeker wordt dan alleen om wille van zijn wetenschappelijkheid binnengehaald. Als er uit het onderzoek geen consequenties worden getrokken, met andere woorden dat - indien dit blijkt - niet publiekelijk ongelijk zal worden bekend, dan is dit onderzoek "negatief maatschappelijk relevant" (p. 124). Hiermee wordt een nieuwe kwalificatie ingevoerd: *maatschappelijke relevantie*. Onderzoek is niet maatschappelijk relevant als er geen conclusies uit worden getrokken, als men dus niet bereid is er (minimaal) zijn ongelijk te bekennen. Hofstee adviseert in dat geval helemaal geen onderzoek te doen.

Samengevat staat hij een onderzoekstype voor, dat moet uitmonden in het gelijk of ongelijk van de deelnemende partijen. Dan alleen is er sprake van een wetenschappelijk evaluatie-onderzoek. Als (een van de) partijen zijn ongelijk niet wenst te erkennen, is het onderzoek niet maatschappelijk relevant en kan dan beter achterwege blijven. Toegepast evaluatie-onderzoek loopt dit risico in sterke mate. Hofstee attendeert er nog op dat dit type onderzoek "... een veel intensiever contact tussen onderzoeker en cliëntsysteem veronderstelt dan gebruikelijk is." (p.125).

De hierboven gegeven definitie van onderzoek lijkt af te wijken van die welke met name in de Amerikaanse methodologie gangbaar was in de jaren 50 en 60. "*The basic aim of science is theory*" schreef Kerlinger (1967). Hij noemde voorspelling en controle ondergeschikte doelen van de wetenschap, hoewel daarin wel de vruchten van de wetenschappelijke arbeid te

plukken zijn: als ik een theorie heb, kan ik ook voorspellen. Of om De Groot te citeren: als ik iets weet, kan ik iets voorspellen; kan ik niets voorspellen, dan weet ik niets (p. 20). Licht bij Kerlinger toch de nadruk op de theorie-vorming, bij De Groot en Hofstee wordt een groter belang gehecht aan de praktische consequentie van het wetenschappelijke werk: het kunnen maken van voorspellingen resp. het kunnen doen van uitspraken wie gelijk of ongelijk heeft.

Samengevat zijn de ideeën van De Groot en Hofstee over onderzoek gelijklopend: (goed) onderzoek moet voorspellingen en/of conclusies opleveren. Het moet volgens vaste regels zijn verricht (methodologisch model: bijv. empirische cyclus of weddenschapsmodel), waardoor het wetenschappelijke waarde krijgt en de resultaten moeten worden gebruikt op een eigenlijke manier, waardoor het maatschappelijk relevant wordt. Zowel De Groot als Hofstee noemen evaluatie-onderzoek het prototype van het onderzoeksmodel. Hiermee is genoegzaam aangetoond dat (programma-) evaluatie, een specifieke vorm van evaluatie-onderzoek, gerekend moet worden tot de grondvorm van (wetenschappelijk) onderzoek, indien het aan de genoemde kwalificaties voldoet.

2.3 HET VERSCHIL TUSSEN ONDERZOEK EN EVALUATIE.

In de vorige paragraaf is duidelijk gemaakt, dat er in methodologisch opzicht geen onderscheid is tussen (wetenschappelijk) onderzoek en evaluatie-onderzoek. Diverse auteurs noemen evaluatie-onderzoek zelfs het basismodel voor onderzoek. Echter, sinds de jaren zeventig heeft evaluatie-onderzoek een bijzondere betekenis gekregen. De nadruk verschuift van *onderzoek* naar *evaluatie*. Was voorheen ieder onderzoek een evaluatie van hypothesen, en kon daarom van evaluatie- of evaluatief onderzoek gesproken worden, nu was het doel *evalueren* dat op een verantwoorde manier (wetenschappelijk, dus onderzoeksmatig) diende te geschieden. Er lijkt nauwelijks een verschil tussen dit evaluatie-onderzoek en het zgn. evaluatieprobleem zoals De Groot (1961) het noemt.

In het psychotherapie-onderzoek met name heeft dit evaluatieprobleem grote aandacht gekregen. Nagegaan werd bijv. welke benadering effectiever was: gedragstherapie of "client-centered" therapie (Sloane e.a., 1975), of welke methode de grootste gedragsverandering veroorzaakte: desensitisatie of flooding (Marks, 1982). De vergelijking werd ook wel gemaakt tussen wel en geen beïnvloeding (experimentele versus controle groepen). Tal van ontwerpen werden voor deze onderzoeken gemaakt: "quasi-experimental designs" genoemd (Campbell & Stanley, 1963; Cook & Campbell, 1979). Deze werden aanvankelijk ontwikkeld voor onderzoek in het onderwijs, maar zijn ook goed bruikbaar gebleken voor het psychotherapie-onderzoek. Deze ontwerpen heten "pre- of

quasi-experimenteel" omdat in deze gebieden geen sprake kan zijn van een echt experiment "in which an experimenter having complete mastery, can schedule treatments and measurements for optimal statistical efficiency, with complexity of design emerging only from the goal of efficiency". (Campbell & Stanley, 1963, p. 171). Het eerste pre-experimentele design is het zgn. "one-shot case study": één groep wordt slechts één keer onderzocht wat inhoudt dat na een behandeling gekeken wordt naar het effect ervan.

Campbell en Stanley zeggen dat dit design toentertijd het meest gebruikt werd in onderwijs-onderzoek, kennelijk de term onderzoek breed definiërend. Voor het psychotherapie-onderzoek zou toen en nu hetzelfde gezegd kunnen worden. Ook wanneer gecompliceerde modellen werden opgesteld met bijv. zgn. wachtlijstgroepen in tal van gradaties of zelfs "double blind-designs" (Wojciechowski, 1984) leidde dit nauwelijks tot generaliseerbare onderzoeksresultaten, een belangrijk doel van dit experimentele onderzoek. De behandeling bleek steeds enigszins effectief; de verschillen tussen de methoden waren vaak slechts gering en mogelijkwijze eerder aan tal van andere factoren te wijten dan aan de onderzochte beïnvloedingsmethode(n).

Al deze pre-, quasi of echte experimentele designs waren onderzoeksmodellen die vooral gericht waren op *nomothetische* kennisvermeerdering, de primaire doelstelling van onderzoek (Kerlinger, 1967). Bij de eerdergenoemde accentverschuiving van onderzoek naar evaluatie, verschoof de doelstelling van *nomothetisch* naar *ideografisch*. Het ging er minder om (wetenschappelijke) kennis te verwerven dan wel mogelijk kunnen aan te tonen. Er is nu sprake van een evaluatie van *doelstellingen*: wordt bereikt wat men beoogde te bereiken. Hoewel De Groot deze vorm van onderzoek reeds in 1961 beschreef, was deze niet eerder zo in de belangstelling dan sinds medio 1970.

Het onderscheid tussen deze twee onderzoekstypen wordt door Anderson & Ball (1978) en Stokking (1984) aangeduid als onderzoek dat gericht is op het verwerven van kennis en onderzoek gericht op het oplossen van problemen. Stokking constateert dat dit onderscheid methodologisch gesproken niet zo scherp is (p. 11). In het verdere betoog zal worden aangetoond waarom deze conclusie hier niet wordt onderschreven, althans niet voor alle vormen van evaluatie-onderzoek, en wat de consequentie is van dit *wel* bestaande methodologische verschil. Stokking (1984) omschrijft evaluatie-onderzoek als een vorm van maatschappelijk gericht onderzoek, waarbij het gaat "*om het beoordeelbaar maken van effecten van interventies, doelgerichte pogingen tot ingrijpen in de werkelijkheid*" (p. 11). In hoeverre dit onderzoek wetenschappelijke waarde kan worden toegekend is volgens hem afhankelijk "*van de hoeveelheid theorie die er wordt ingestopt*" (p. 17), hiermee in feite Kerlinger volgend. Hij is wel van mening dat "theorie" breed moet worden opgevat, het mag om impliciete theorie gaan of zelfs om vooroordelen en de resultaten zullen

een gebrek kunnen hebben aan generaliseerbaarheid. Er zal altijd sprake zijn van "kenniswinst". Als kenniswinst het belangrijkste element van onderzoek is, wordt de definitie wel heel breed en bestaat er inderdaad weinig onderscheid meer tussen de diverse vormen van onderzoek voor wat betreft hun wetenschappelijke waarde. Als de wetenschappelijke waarde evenwel wordt afgemeten aan de hoeveelheid theorie die in het onderzoek is gestopt, ontstaan er dus twee vormen van onderzoek, wetenschappelijk en niet-wetenschappelijk, die beide wel kennis kunnen opleveren.

Het door Stokking beschreven conflict tussen evaluatietheoretici die zich afkeren van de gebruikelijke sociaal wetenschappelijke methoden en de sociale wetenschappers die zich te weinig met evaluatie-onderzoek bezighouden zal menige evaluator aan den lijve ervaren hebben: Evaluatie-onderzoek in de praktijk wordt door veel sociale wetenschappers als "onwetenschappelijk" en dus als "slecht" of minstens "zacht" gekwalificeerd. Ook Stokking betreurt dit, hoewel hij dit conflict in feite bekrachtigt.

Als noch bij De Groot noch bij Hofstee argumenten voor deze diskwalificatie worden gevonden, wat is dan de achtergrond voor deze opvatting? Welke eigenschappen worden er aan evaluatie-onderzoek toegekend, of liever, wat mist evaluatie dat hieraan door velen de kwalificatie "*wetenschappelijk*" wordt onthouden? Enkele in het oog springende kenmerken zijn: het ontbreken van een experimenteel-manipulerend karakter, het gemis aan controle of vergelijkende groepen en het niet expliciet aanwezig zijn van een theoretisch kader. Volgens bovenstaande citaten van o.a. De Groot en Hofstee zijn dit geen onvoorwaardelijke condities voor wetenschappelijkheid. Daarnaast is tal van veldonderzoek eveneens niet experimenteel en bevat veel survey-onderzoek ook geen controlegroepen. Een theorie is (impliciet) wel aanwezig, al was het maar dat twee conflicterende vraagstellingen op houdbaarheid worden onderzocht: wie heeft gelijk, heeft een bepaalde werkwijze (behandeling) wel of geen effect, waarbij beide partijen een theorie of hypothese hebben waarom ze gelijk menen te hebben.

Een, ook door Stokking genoemd, belangrijk gemis van evaluatie-onderzoek in een specifieke praktijksituatie is het gebrek aan *generaliseerbaarheid* van de resultaten. De conclusies betreffende de effectiviteit van een bepaalde werkwijze van een welomschreven G.G.Z.-instelling zijn niet geldig voor een naburige instelling. De verschillen tussen belangrijke variabelen zoals patienten- en therapeutkenmerken, behandelingsmethode, e.d. zullen daarvoor doorgaans te groot zijn. Een ander bezwaar is het descriptieve karakter van genoemd onderzoek. De Groot en Hofstee volstaan met de doelstelling van onderzoek dat, op basis van de resultaten, voorspellingen gedaan kunnen worden of dat aangegeven kan worden wie gelijk heeft. Evenwel betekent het gemis aan een theorie dat doorgaans geen verklaring voorhanden is voor het gevonden verschijnsel, zoals Stokking beweert. "Wetenschap bedrijven is evenwel niet alleen weten wanneer je 'goed zit'

maar ook waarom". (Stokking, p. 20, p. 31). Evaluatie-onderzoek zal het antwoord op deze waarom-vraag niet of slechts ten dele kunnen geven. Maar hiermee is het nog niet a-wetenschappelijk zoals inmiddels is betoogd.

Evaluatie-onderzoek is kennelijk een zeer breed begrip waar de diverse auteurs verschillende kenmerken aan toeschrijven of eisen aan stellen al naar gelang meer het accent ligt op *onderzoek* dan wel op *evaluatie* en bijgevolg de doelstelling meer betrekking heeft op *kennen* dan wel op *kunnen*. (zie ook Swanborn, 1981, p. 113). Er bestaat een Babylonische spraakverwarring tussen de diverse theoretici wat nu precies onderzoek, evaluatie en evaluatie-onderzoek genoemd moet worden en wat de verschillen tussen deze begrippen zijn. Hier wordt geprobeerd deze begrippen opnieuw te omschrijven en ze ook methodologisch van elkaar te onderscheiden. Het hele scala van handelen voltrekt zich tussen zuiver onderzoek enerzijds en zuiver handelen anderzijds. De tussenstadia tussen deze twee uitersten worden door Barlow e.a. (1984) aangeduid als "*treatment research*" en "*treatment evaluation*", twee pijlers in de spreekwoordelijke kloof tussen onderzoek en praktijk. In tabel 2.1 worden deze vier begrippen met hun primaire en secundaire doelen aangeduid. Met evaluatie-onderzoek wordt nu zowel "*treatment research*" als "*treatment evaluation*" aangeduid. De omkering van het primaire en secundaire doel heeft echter belangrijke methodologische consequenties. (zie ook Sijben, 1984a).

Tabel 2.1 Het onderscheid tussen onderzoek en behandeling en de tussenstadia.

Doel:	Zuiver onder- zoek	Behand. onder- zoek	Behand. evalu- atie	Zuivere behandeling
betere wetenschap- pelijke uitspraken	primair	primair	secundair	niet
betere behande- lings- resultaten	niet	secundair	primair	primair

(naar: Barlow, 1984, The Scientist-practitioner., p. 284)

Bij (behandelings)onderzoek is er primair sprake van een *theoretisch kader* waaruit een of meer vraagstellingen worden afgeleid, welke in een zo goed mogelijk gespecificeerd en

gestandaardiseerd onderzoeksontwerp worden getoetst. De standaardisering houdt in, dat vooraf wordt vastgelegd welke personen met welke kenmerken aan welke behandelingscondities worden onderworpen. De experimentele conditie zoals die in het onderzoeksontwerp met zijn specificaties is vastgelegd, is de *onafhankelijke variabele*, welke door de onderzoeker op basis van zijn theorie en vraagstelling wordt bepaald, zo mogelijk gemanipuleerd. De onderzoeksresultaten zijn de *afhankelijke variabelen*. Volgens *statistische toetsing* wordt nagegaan of de gevonden verschillen (bijv. tussen voor- en nameting of tussen de verschillende onderzoekscondities) voldoende groot zijn om hier betekenis aan toe te kennen (*statistische significantie*). Er is sprake van *normatief* of *relatief* meten (Warries, 1970). Een bijkomend kenmerk van behandelingsonderzoek is, dat dit zich doorgaans afspeelt in een laboratorium-achtige setting en dat de resultaten via *wetenschappelijke publikaties* bekend gemaakt worden, hetgeen inhoudt dat deze slechts terloops of via een grote omweg (opleiding) de praktijk kunnen beïnvloeden.

Behandelingevaluatie verschilt in belangrijke onderdelen van het hierboven beschreven proces. Deze is primair *praktijkgericht*: de bestaande praktijksituatie is uitgangspunt. Een of meer aspecten van het functioneren dienen geëvalueerd te worden. Deze zullen in eerste instantie op de effectiviteit van de gegeven behandeling betrekking hebben. Maar ook hieraan gerelateerde variabelen zoals cliëntvariabelen en organisatorische procedures kunnen onderdeel uitmaken van de evaluatie. Het eerste oogmerk ervan is nagaan of de behandeling aan zijn *doel* beantwoordt. Deze doelen dienen daarom vooraf gespecificeerd te worden. Doorgaans zijn deze slechts impliciet aanwezig. Het behandelingsproces is erop gericht, met alle mogelijke middelen, de doelen te bereiken. De variaties waar de behandelaar o.a. gebruik van kan maken is de tijdsduur die voor de totale behandeling wordt gebruikt, de frequentie en de intensiteit ervan en de gehanteerde methode en strategie. Het behandelingsproces is bijgevolg eerder een *afhankelijke variabele* en het behandelingsresultaat de *onafhankelijke*. De toetsing van dit resultaat dient derhalve niet normatief maar *criterium-georiënteerd* te zijn (Van der Linden, 1980). Het verschil tussen statistische significantie en klinische relevantie (Meehl, 1954) speelt ook hier een rol: De laatste is belangrijker dan de eerste. De *feedback* van de resultaten zal snel zijn: rechtstreeks aan betrokkenen, het gaat immers over hun eigen functioneren. De resultaten kunnen gebruikt worden om dit te verbeteren. Het nadeel kan zijn, dat meer wordt aangetoond *hoe* het functioneren is dan *waarom* dit zo is. Dit is afhankelijk van de uitgebreidheid van de verzamelde gegevens en de mogelijke relevantie ervan voor het behandelingsresultaat. Maar dit geldt voor iedere vorm van (evaluatie-)onderzoek.

In tabel 2.2 worden deze twee vormen van evaluatieonderzoek nog eens op hoofdpunten tegenover elkaar gezet. (Sijben, 1980b; zie ook Veerman, 1983, p. 231). Deze ver-

schillen zijn niet steeds even duidelijk of absoluut, maar geven meer de accentverschillen aan. Duidelijk zal zijn, dat het scharen van twee zo uiteenlopende begrippen onder een noemer wel tot misverstanden aanleiding moet geven. De kloof waar Barlow van sprak tussen onderzoek en praktijk wordt versmald door het invoeren van behandelingsonderzoek en behandelingsevaluatie. Tussen deze twee blijft evenwel toch nog een kloof bestaan. Of deze overbrugbaar is, is de vraag. Of dat noodzakelijk is, eveneens. Elk heeft zijn eigen doelstellingen, werkwijzen en functies. Het zou wel wenselijk zijn als er inhoudelijk meer raakpunten zouden zijn. Goed opgezette en uitgevoerde behandelingsevaluatie zou hier een belangrijke bijdrage aan kunnen leveren. De meer theoretisch gefundeerde waarom-vragen zouden door het *behandelingsonderzoek* beantwoord kunnen worden, maar dit is hier tot op heden nauwelijks in geslaagd (zie ook hfst. 4). Een goed opgezette *behandelingsevaluatie* is hiertoe wellicht meer in staat. Door zijn nauwe aansluiting bij de praktijk zou dit een grote kans van slagen kunnen hebben.

Tabel 2.2 Verschillen tussen behandelingsonderzoek en behandelingsevaluatie.

ONDERZOEK	<-----TYPE----->	EVALUATIE
toetsen van theorieën	<-----doel----->	toetsen van effecten
experimenteel	<-----plaats----->	praktijksituatie
(quasi)experimenteel design	<-----methode----->	bestaand behandelings methode
gestandaardiseerd	<-----proces----->	geïndividualiseerd
onderzoeksproces	<onafhankelijke-variabele>	beoogd behandelings resultaat
resultaat (verschil voor-nameting)	<--afhankelijke-variabele-->	behandelingsproces
normatief/relatief	<----toetsing---->	absoluut/criterium georiënteerd
statistische significantie	<---waardering--->	klinische relevantie
publicatie (lang)	<----feedback----->	aan betrokkenen
conclusie gericht	<---resultaat---->	beslissingsgericht

Hiermee zijn deze drie begrippen: (zuiver) onderzoek evaluatie-onderzoek en (programma-)evaluatie gedefiniëerd. Evaluatie-onderzoek bestrijkt zowel (zuiver) onderzoek als (programma-)evaluatie. Het zuivere of behandelingsonderzoek heeft een meer experimenteel karakter, terwijl behandelings- of programma-evaluatie praktijkgericht is. Methodologisch is

er een belangrijk onderscheid: de afhankelijke en onafhankelijke variabelen zijn verwisseld. De wetenschappelijkheid van deze verschillende onderzoeksvormen wordt niet bepaald door hun hoedanigheid maar door hun kwaliteit.

Aldus is de plaatsbepaling van programma-evaluatie als vorm van behandelingsevaluatie aangegeven. Tenslotte worden in de volgende paragraaf nog enkele vormen van deze evaluatie beschreven.

2.4 VORMEN VAN BEHANDELINGSEVALUATIE.

2.4.1 Summatieve of formatieve evaluatie

Een terrein waar de behandelingsevaluatie of beter gezegd: evaluatie van het handelen in zijn algemeenheid reeds vele jaren wordt toegepast, is het onderwijs. Diverse vormen van evaluatie zijn hier in de loop der jaren ontwikkeld, elk met hun eigen bedoelingen. De meest bekende is het afsluitende examen, maar sinds de jaren zestig worden nieuwe, tussentijdse evaluatievormen gebruikt. Gelet op de inhoudelijke overeenkomst tussen het onderwijs aan leerlingen en de behandeling van patiënten - beiden dienen gedurende een zekere periode door een "expert" zich iets eigen te maken, kennis en/of handelingsbekwaamheid, die zij voordien nog niet bezaten -, kunnen de evaluatiemodellen in het onderwijs mutatis mutandis in de behandelingssituatie van de G.G.Z. eveneens bruikbaar zijn. Deze analogie wordt hier kort geschetst. Sinds het onderwijs is geïnstitutionaliseerd, is in toenemende mate de aard, inrichting, omvang qua lesuren, leerstof, e.d. vastgelegd tot en met het eindresultaat in exameneisen. De beoordeling van dit eindresultaat was normatief: de prestaties van een individuele leerling werden beoordeeld in relatie tot die van de overige leerlingen of van een vergelijkbare groep, bijv. een ander leerjaar of een landelijke normgroep (bijv. Cito basistoets). Het cijferen geschiedde volgens de "wet van Posthumus": de verdeling van de uitslagen vertoonde een normale verdeling welke in 10 delen werd gesplitst met de modus bij de 6. Deze werkwijze leidde tot veel klachten en problemen: standaardisering van het leerproces was voor sommige leerlingen zeer nadelig, het lestempo ging te vlug ("de leraar heeft het boek al uit, maar ik nog niet"), de vele onvoldoendes betekenden een inefficiëntie van het schoolprogramma en de eenmalige eindtoets, ook wel *summatieve toetsing* genoemd, gaf te laat informatie om als feedback te kunnen dienen. Eigenschappen van summatieve evaluatie zijn, dat de norm achteraf wordt bepaald, relatief is en er wordt slechts eenmaal getoetst.

Om het onderwijsrendement te optimaliseren, om met name die leerlingen er het meest van te laten profiteren die hier de grootste behoefte aan hebben, gelet op hun achterstand,

intelligentie, motivatie, etc., werd het onderwijs geïndividualiseerd. Er werd gestreefd naar het bereiken van een *maximale prestatie* voor iedere leerling door de tijd, aandacht, leerstof en lestempo zo goed mogelijk af te stemmen op de individuele leerling. Deze individualisering betekende dat heel precies omschreven werd wát de leerlingen dienden te weten/kennen/kunnen aan het einde van het leerproces: de zgn. *leerdoelen*. Voorheen waren deze doelen impliciet in het leerboek aanwezig. Bij deze nieuwe werkwijze werd van dit boek vaak afgeweken; er werden o.a. nieuwe didactische vormen toegepast, zodat zoveel mogelijk leerlingen de leerdoelen zouden kunnen bereiken. Welke doelen dit waren werd nu vooraf geëxpliciteerd.

Om tussentijds te kunnen bewaken of de individuele leerlingen "*doelgericht*" werkten, was het nodig de leerdoelen regelmatig te toetsen; dit gebeurde met zogenaamde diagnostische of formatieve toetsen. Eventuele hiaten in de kennis konden dan worden aangetoond en deze feedback diende om middels herhalingsprogramma's, extra lesstof of huiswerk, deze zonodig op te vullen. De functie van de evaluatie veranderde van louter constaterende en normatief beoordelende, summatieve evaluatie, in op feedback gerichte, *formatieve* evaluatie, gerelateerd aan vooraf gespecificeerde criteria of doelstellingen. Evaluatie diende nu ook om de *prestaties zelf* te verbeteren. Bloom e.a. (1971) vat de functies van dergelijke evaluatie in het onderwijs als volgt samen:

1. Evaluation as a method of acquiring and processing the evidence needed to improve the student's learning and the teaching.
2. Evaluation as including a great variety of evidence beyond the usual final paper and pencil examination.
3. Evaluation as an aid in clarifying the significant goals and objective of education and as a process for determining the extent to which students are developing in these desired ways.
4. Evaluation as a system of quality control in which it may be determined at each step in the teaching-learning process whether the process is effective or not, and if not, what changes must be made to ensure its effectiveness before it is too late.
5. Finally, evaluation as a tool in education practice for ascertaining whether alternative procedures are equally effective or not in achieving a set of educational ends.

Handbook on formative and summative evaluation of student learning. Bloom e.a. (1971) p. 7-8

In bovenstaande is "*education*" te vervangen door "*mental health*", "*student*" door "*client*" en "*teacher*" door "*therapist*". Het "*teaching-learning-process*" is vergelijkbaar met

het *hulpverleningsproces* in de geestelijke gezondheidszorg. De bovenbeschreven evaluatie-oogmerken kunnen ook hier onderscheiden worden: evaluatie van processen en resultaten, vaststellen van doelen en evalueren ervan, kwaliteitscontrole en nagaan of andere behandelingsvormen wellicht efficiënter en/of effectiever werken. De evaluatie kan summatief zijn, d.w.z. op eindtermen zoals een (gemiddelde) score op een standaard vragenlijst. Een voorbeeld van een dergelijke evaluatie is de Denver Community Mental Health Questionnaire (Ciario e.a., 1976, 1978, 1981; Sijben, 1979a). In Nederland wordt een standaardklachtenlijst of een (psychiatrisch) interview zoals de SCL of PSE (Hodiamont, 1986) hiervoor wel gebruikt. Een goed voorbeeld van een doelgeoriënteerde evaluatiemeting, welke ook als eindevaluatie gebruikt wordt, is de Goal Attainment Scaling (Kiresuk & Sherman, 1968; Melief e.a., 1979) (zie ook hoofdstuk 4). Belangrijker is ook hier de formatieve evaluatie die tussentijds plaatsvindt. Deze is bedoeld om het behandelingsplan bij te stellen, wanneer de doelen niet gerealiseerd dreigen te worden.

Beide evaluatievormen, formatief en summatief, hebben elk hun eigen bedoelingen. Uit de beschrijving zal duidelijk zijn, dat in een situatie als de G.G.Z., waarin doorgaans sprake is van een individuele benadering, met per patiënt geldende doelen en grote nadruk op optimaliseren van het handelen zelf, terwijl de behandelingsduur flexibel is, formatieve evaluatie de voorkeur verdient.

2.4.2 "Process" of "outcome" evaluatie.

In het psychotherapie-onderzoek wordt een onderscheid gemaakt tussen *proces-evaluatie* en *effect-evaluatie*. De doelen van deze twee onderzoekstypen zijn verschillend: effect-evaluatie richt zich met name op het resultaat van de behandeling of liever het verschil tussen het begin en het einde. Proces-evaluatie handelt "over wat er tussen begin en einde gebeurt" (Soudijn, 1982, p. 66). Het doel is na te gaan waarom of waardoor de effecten ontstaan. Hiertoe worden bijv. alle gesprekken op band vastgelegd en later geanalyseerd om na te gaan welke interventies tot welk resultaat hebben geleid.

Aan beide vormen zijn nadelen te onderkennen. Bij effect-evaluatie is de keuze van de effectmaten een probleem (zie hoofdstuk 4), bij proces-evaluatie levert de inhoudsanalyse van het verzamelde bandmateriaal veel interpretatiemoeilijkheden op. Op een conferentie over de dialoog tussen onderzoek en praktijk in de psychotherapie kwalificeerde Soudijn het procesonderzoek als het kijken naar een videoband van een goochelaar vanuit de positie van de toeschouwer: je ziet niet wat je eigenlijk zou willen zien, nl. wat er achter het uiterlijk van zijn kunsten schuilgaat (Soudijn, 1984). Zo zie je bij een therapeut evenmin de bedoelingen van zijn verbale interventies. Het duidelijkst manifesteert zich dit bij een

paradoxale benadering: de inhoud en de bedoeling van de interventie zijn hier tegengesteld.

Effect-evaluatie lijkt dan een methodiek die meer aansluit bij de behandelingssituatie. Voor de praktijksetting zelf zal procesonderzoek nauwelijks kunnen renderen: directe resultaten voor het behandelingsproces of -evaluatie mogen niet worden verwacht. Voor het aantonen van causale verbanden zijn bij voorkeur experimenten aangewezen, "... *but clinical psychotherapy has proven a peculiarly difficult subject for true experiments*" (Orlinsky and Howard, 1978, p. 319-320). Daarnaast geldt dat, wil de relatie tussen interventie en effect aangetoond kunnen worden, eerst dit effect zelf meetbaar moet zijn. Hierop wordt in hfst. 4 nader ingegaan.

Wellicht gelet op de nadelen van beide evaluatiemethoden, hebben deze twee vormen van psychotherapie-onderzoek in een wisselende mate van belangstelling gestaan. Effectstudies hebben evenwel aanzienlijk meer aandacht gekregen dan procesonderzoek. Toch hebben al deze effectstudies weinig resultaten opgeleverd voor de behandelingssituatie. Dit zou mede veroorzaakt kunnen worden door het feit dat de meeste studies eerder *behandelingsonderzoek* dan *behandelingsevaluatie* zijn. De zo noodzakelijke *generalisatie* kan niet worden gemaakt door het grote verschil tussen onderzoeks- en praktijksituatie. Een extra bezwaar is dat de meeste effectstudies zich nauwelijks met "echte" patiënten hebben beziggehouden. Studenten psychologie waren doorgaans de (verplichte) proefpersonen en ook dikwijls de behandelaars.

In een meta-analytisch overzicht van 143 effectstudies melden Shapiro en Shapiro (1983): "*It is concluded that this meta-analysis underlines the urgent need for greater methodological diversity and clinical realism in therapy research*". In toenemende mate ontstond er behoefte aan effectstudies in natuurlijke behandelingssituaties. Deze behoefte was zowel *intrinsiek* bij de behandelaars en hun cliënten, als *extrinsiek* bij de leiding, directie en financierende instanties. Zo ontstond een nieuwe evaluatiemethodiek in de geestelijke gezondheidszorg die aansloot bij deze praktijksituatie, welke in het onderwijs reeds langer functioneerde als formatieve evaluatie.

2.4.3 Programma-evaluatie.

De aan het slot van de vorige paragraaf genoemde evaluatievorm is programma-evaluatie. Deze kan gezien worden als een speciale vorm van behandelingsevaluatie, waarbij nu een heel programma het onderwerp is. Dit programma kan één cliënt zijn, een groep van cliënten, een groep van behandelaars of een hele instelling. Het is een vorm van effectonderzoek met een sterk formatieve inslag. Programma-evaluatie voldoet aan de definitie van De Groot voor empirisch toetsend onderzoek als aan de eisen van Hofstee ten aanzien van het weddenschapmodel; met behulp van programma-evaluatie moet

worden bepaald wie *gelijk* heeft bij de beantwoording van de vraag of een behandeling effect heeft of niet.

Betoogd is dat programma-evaluatie geheel kan voldoen aan wetenschappelijke kwalificaties, althans dat de onderzoeksmethodiek dit niet uitsluit. De discussie tussen de sociale wetenschappers hierover is onterecht en zinloos ("*... worrying about whether evaluation is or is not research seems to be particularly unproductive way for evaluators to spend their time.*" Anderson en Ball, 1978, p. 11). Alleen de wijze waarop het onderzoek is uitgevoerd, de objectiviteit, de betrouwbaarheid en validiteit van het gebruikte instrumentarium bepalen de wetenschappelijke waarde, zoals dit voor alle onderzoek geldt. Daaraan moet volgens Hofstee nog de *maatschappelijke relevantie* worden toegevoegd. Hofstee is sceptisch ten aanzien van het verrichten van onderzoek in de praktijk, omdat hij vreest dat de partijen geen *ongelijk* zullen erkennen en het onderzoek slechts als "public relations" wensen aan te wenden of om tijd te winnen, hetgeen het maatschappelijk irrelevant maakt.

Ondanks deze sombere voorspelling wordt met programma-evaluatie getracht in een praktijksituatie de effectiviteit van het functioneren te bepalen. De bedoeling is, zoals dit formatieve toetsing betaamt, om deze als feedback te gebruiken. Of Hofstee gelijk heeft of niet zal de toekomst moeten leren. Programma-evaluatoren hopen van niet, en om meerdere redenen dan hun eigen broodwinning.

Hoofdstuk 3:

Programma-evaluatie, wat is dat?

*Program evaluation is a proces
by which society learns about itself.*
(Cronbach, 1e stelling, 1980)

3.1 INLEIDING.

Zoals in hoofdstuk 2 is beschreven, bestaan er duidelijke verschillen tussen *onderzoek* gericht op (wetenschappelijke) kennisvermeerdering en *evaluatie* van het functioneren van een bepaalde werkwijze of instelling. Maar al te dikwijls worden én de onderzoeker én de werkers in de praktijk gefrustreerd als geprobeerd wordt een wetenschappelijk onderzoek in het veld uit te voeren. De onderzoeker wil de praktijksituatie aanpassen aan zijn onderzoeksvraagstelling, terwijl de hulpverlener daarentegen wil weten of datgene wat hij doet goed is, in overeenstemming is met datgene wat hij geacht wordt of wenst te doen. Voor dit doel is programma-evaluatie een aangewezen procedure.

Wat onder programma-evaluatie wordt verstaan, zijn herkomst en verschijningsvormen, wordt in dit hoofdstuk beschreven. Tenslotte wordt in hoofdlijnen het verloop van een programma-evaluatie-project geschetst, de informatie die dit kan opleveren en hoe deze als feedback kan dienen voor de diverse geledingen van een instelling: patiënt, therapeut, team, leiding, bestuur en de externe, controlerende en financierende instanties. Ook de rol van de evaluator die met zijn boodschap veel beroering teweeg kan brengen, komt aan bod.

3.2 HISTORIE EN DEFINITIE.

Zoals in het inleidende hoofdstuk reeds is genoemd en de naam overigens ook doet vermoeden, stammen de ideeën over programma-evaluatie uit de U.S.A. Begin jaren zestig ontstond aldaar een grote behoefte om de invloed van de federale overheid te vergroten; enerzijds om de bestaande overheidsdiensten op het gebied van onderwijs en geestelijke gezondheidszorg beter te controleren en anderzijds om nieuwe ontwikkelingen te kunnen sturen (Cronbach, 1980).

Grote "nationale" programma's als "Head Start" werden opgezet, gericht op bepaalde bevolkingsgroepen in achterstandsituaties. Evaluatie van deze programma's was een nood-

zaak bij de toekenning van de subsidies. Aanvankelijk onder de verzamelnaam "*Evaluation Research*" kwam een evaluatiemethodologie weer in de aandacht die hierbij aansloot: praktijkgericht en ter evaluatie van de geformuleerde doelstellingen, de effectiviteit van het programma. Deze methodologie was niet nieuw; zo citeert Suchman (1967) een publicatie uit 1952, waarin Riecken evaluatie definieert als "*the measurement of desirable and undesirable consequences of an action that has been taken in order to forward some goal that we value*" (p. 28). Het onderscheid tussen deze evaluatievorm en wetenschappelijk onderzoek wordt in 1955 door het Subcommittee on Evaluation of Mental Health Activities reeds geformuleerd: "*Evaluation thus connotes scientific method, but has characteristics that distinguish it from that type in order to formulate hypotheses and theory for the sake of new knowledge itself, irrespective of judgment of the value of the knowlegde*" (Suchman, p. 29).

De term "*program evaluation*" was ook reeds in 1955 gebruikt ter evaluatie van "*Public Health Programs*". De toen geformuleerde definitie is nog steeds geldig: "*Evaluation of public health programs should denote a measurement of the effectiveness of the program. This effectiveness should be measured in terms of the fulfillment of the program's objectives*" (Greenberg en Mattison, 1955). Later wordt deze omschrijving ingeperkt tot "*Program evaluation can be defined as the measurement of success in reaching a stated objective*" (James, 1962). Het begrip "*Program*" wordt gedefinieerd als "*A sponsered activity more after then not from public funds, aimed at mitigating a social or economic welfare*" (Perloff, 1976). Worden deze begrippen naar de Nederlandse situatie vertaald, dan kan programma-evaluatie gedefinieerd worden als: *het vaststellen van de waarde van een programma door de effecten ervan te vergelijken met de doelen die voor dit programma zijn gesteld* (Sijben, 1984a). Het inmiddels ingeburgerde woord "*programma*" kan naar analogie van Perloff voor Nederland betrekking hebben op het functioneren van de (geestelijke) gezondheidszorg, sociale voorzieningen en het onderwijs.

Hiermee is het belangrijkste doel van programma-evaluatie aangegeven: het leveren van *feedback* aan een instantie over het functioneren van een bepaald onderdeel. Deze doelstelling is descriptief: de evaluator legt de resultaten op tafel die de effectiviteit aangeven en laat het verder aan de opdrachtgever over ermee te doen wat hem goeddunkt. Bij deze strikte opvatting is een kanttekening te maken: moet de taak van de evaluator zich hiertoe beperken? Veel programma-evaluatie-projecten gaan inderdaad niet verder dan het geven van een beschrijving van de status quo. De resultaten zijn een spiegel die de organisatie wordt voorgehouden: mogelijke tekortkomingen wordenesignaleerd. Over het *waarom* ervan wordt geen uitsluitel gegeven. Een voorbeeld van een dergelijk project is de Evaluatie Krisis Centrum Utrecht (Beenackers en Donker, 1981). Oorzaken van dysfunctioneren kunnen misschien worden afgeleid uit de gegevens of anders zal

middels specifiek onderzoek naar de oorzaken van het gebrek aan effectiviteit kunnen worden gezocht.

Niet iedereen hangt deze beperkte opvatting aan. Zo geeft Suchman (1967) verschillende ideeën weer over een evaluatieproject (p. 30 e.v.).

- Klineberg is van mening dat *"evaluation is a process which enables the administrator to describe the effects of his program, and thereby to make progressive adjustments in order to reach his goal more effectively"*.

- Bigman gaat verder; als tweede en derde hoofddoel van evaluatie, naast het evalueren van het bereiken van de doelen, noemt hij: *"to determine the reasons for specific successes and failures"* en *"to uncover the principles underlying a successful program"*. Hiermee komt programma-evaluatie sterk op het terrein van het vergelijkende onderzoek enerzijds, maar begeeft zich ook op het gebied van het beleid van het programma. Als zesde doel noemt hij dan ook: *"to redefine the means to be used for attaining objectives, and even to redefine subgoals, in the light for research findings"*. Of een dergelijke brede, diepgaande en verdragende taakstelling praktisch haalbaar is, kan worden betwijfeld. Eenduidige antwoorden op de gestelde waarom-vraag zijn zeker niet zodanig te geven, dat deze direct vertaald kunnen worden in beleid. Daarnaast krijgt de evaluator een veelzijdige en machtige rol toebedeeld: hij weet *wat* de beste werkwijze is, *waarom* dit zo is en moet daarom het nieuwe beleid bepalen, alsof beleid uitsluitend bepaald zal of kan worden door effectiviteit. Vooralsnog zal de evaluator zich veel bescheidener moeten opstellen.

Suchman geeft het evaluatieproces met de volgende zes stappen aan:

1. Indication of the goals to be evaluated.
2. Analysis of the problems with which the activity must cope.
3. Description and standardization of the activity.
4. Measurement of the degree of change that take place.
5. Determination of whether the observed change is due to the activity or to some cause.
6. Some indication of durability of the effect.

(Suchman, 1967, p. 31).

Indien evaluatie hieraan voldoet, mag het geplaatst worden *"within the camp of scientific research"*. Stap 3: 'het standaardiseren van een activiteit, maakt dat veel programma-evaluatie projecten hieraan niet zullen voldoen. De praktijk laat zich niet makkelijk standaardiseren, zoals dit voor onderzoeksprojecten juist vereist is. Afgezien hiervan geeft Suchman een bruikbare omschrijving van het programma-evaluatie-proces.

Terugkerend tot de vraag of programma-evaluatie zich in eerste instantie moet beperken tot een *beschrijving* van de status quo, wordt deze hier bevestigend beantwoord met de toevoeging dat alle in de verzamelde gegevens opgesloten informatie welke bij het bepalen van een nieuw beleid behulpzaam kan zijn, natuurlijk moet worden aangewend. Dergelijke informatie is evenwel een *surplus*, het is niet het belangrijkste doel van een programma-evaluatie-project. Dit moet primair antwoord geven op de vraag: *wat is er aan de hand?* Overigens is het geen sinecure dit antwoord op een betrouwbare en valide wijze te geven. Het antwoord op deze vraag dient echter wel gedifferentieerd te zijn. Het functioneren is in diverse doelen onderverdeeld en elk doel afzonderlijk dient geëvalueerd te worden. Tevens kan de algemene informatie per doel worden gesplitst naar subgroepen. Zo kan nagegaan worden of (verminderde) effectiviteit gerelateerd is aan cliëntkenmerken zoals diagnose, demografische gegevens e.d.. of dat er een verband bestaat met behandelings- of behandelaarsvariabelen. Dergelijke bivariate verbanden zijn eenvoudig aan te geven wanneer, zoals te doen gebruikelijk is, een cliëntinformatie systeem c.q. Management Information System (zie par. 3.3.1) gekoppeld wordt aan het evaluatiesysteem. Of dergelijke interrelaties zinvol zijn en informatie verschaffen over de richting van mogelijke beleidsveranderingen, zal mede afhankelijk zijn van de aard en de omvang van dit gegevensbestand.

Al deze informatie is nog steeds *descriptief*, beschrijft diverse aspecten van het functioneren voor verschillende groepen. De voor de hand liggende volgende vraag: *Waarom is dit zo?*, kan soms beantwoord worden met behulp van deze resultaten, maar zal vaak extra onderzoek nodig maken, of zal soms ook niet beantwoordbaar zijn. De "black box" van bijv. de psychotherapie maakt het geven van verklaringen voor bepaalde veranderingen of het ontbreken ervan dikwijls niet mogelijk. Een behandeling is een complex van factoren en invloeden waarvan niet altijd te achterhalen is welke interventie tot welk resultaat heeft geleid (Rutman, 1980, p. 45). Antwoord geven op deze vraag zal een samenspel moeten zijn tussen de evaluator en de opdrachtgever c.q. de medewerkers van het geëvalueerde programma. Zij moeten dan bereid zijn hun eigen functioneren ter discussie te stellen.

De derde vraag: *Wat moet eraan gebeuren?*, is vanzelfsprekend de taak van het management, de beleidsmakers of de financier. De evaluator kan hiervoor net als ieder ander zijn adviezen geven. Was Bigman (zie boven) van mening, dat de evaluator al deze drie vragen zou beantwoorden, hier wordt staande gehouden dat het "slechts" tot de taak van programma-evaluatie kan worden gerekend om een gedifferentieerde beschrijving te geven van de status quo: nagaan of de doelen wel of niet gerealiseerd zijn, met de hierboven beschreven nadere precisering.

Dat dit hier met zoveel nadruk wordt gesteld, vindt zijn reden in het risico dat schuilt in een (te) brede taakopvat-

ting. Wanneer uit een programma-evaluatie-project blijkt, dat bepaalde resultaten niet in overeenstemming zijn met de doelen die vooraf waren gesteld, en de evaluator heeft ook als taak c.q. krijgt deze toebedeeld om het *waarom* hiervan (overtuigend!) aan te tonen, dan zal (kunnen) blijken dat deze vraag niet bevredigend beantwoord kan worden. Immers, achter iedere waarom-vraag ligt een volgende. Als bijv. blijkt, dat het functioneren van een team minder effectief wordt en nadere analyse toont tevens aan, dat dit vooral wordt veroorzaakt door een of twee leden van het team, is een eerste indicatie gegeven waar de oorzaak kan liggen. Vervolgens dringt zich dan de vraag op: waarom bestaan er zulke verschillen tussen deze personen? Doorgaans hebben collegae onderling hierover duidelijke ideeën, maar deze worden niet uitgesproken, ze zijn immers subjectief en kunnen bij hardop verwoorden ervan de groepsgeest verstoren. *"More important than therapy results to any clinic is staff harmony"* schreef Haley (1975). (Zie ook hoofdstuk 11.) De evaluator wordt nu als detective gebruikt om deze verschillen te verklaren. Doorgaans ontbreken hiervoor echter voldoende gegevens. Zo mogelijk worden bijv. nog verschillen gesignaleerd in gebruikte behandelingsmethoden, maar waarom leveren die dan weer verschillen in resultaat op, etc. Wanneer de detective faalt, wordt het probleem niet opgelost, de verklaring niet gevonden, er is "dus" ook niets aan te doen en men gaat over tot de orde van de dag en wacht af of de resultaten zo blijven; of het evaluatie-project wordt gestopt omdat het geen bruikbare verklaringen oplevert. Een dergelijke gang van zaken is niet denkbeeldig en is een (doodlopende) weg door het labyrint van interacties tussen evaluator en geëvalueerde. Een duidelijke en beperkte taak van het evaluatieproject is een eerste vereiste om dit doolhof te vermijden en de rol van de evaluator voor alle betrokkenen (inclusief voor hemzelf) zo duidelijk mogelijk te maken. Over deze rol van de evaluator, de gebruikswaarde van de evaluatiegegevens en hun feedback-functie zal in par. 3.5 nader worden ingegaan.

3.3 VORMEN VAN PROGRAMMA-EVALUATIE.

Programma-evaluatie is wel eens een *"container begrip"* genoemd (Schnabel, 1984b) waar vele activiteiten en evaluatiemethodieken in kunnen worden ondergebracht. In deze paragraaf wordt deze container van compartimenten voorzien. In een tweedeling wordt eerst onderscheid gemaakt in een *inhoudelijke* en *organisatorische* afdeling. In par. 3.3.1 worden de verschillende evaluatievragen, de aandachtsvelden van programma-evaluatie beschreven, terwijl in par. 3.3.2 de diverse organisatievormen van programma-evaluatie aan de orde komen.

3.3.1 De evaluatievragen.

De gegeven definitie van programma-evaluatie en de hier bepleitte beperkte taakstelling laten nog veel ruimte over voor een concrete invulling van de evaluatievragen. Onder effectiviteit van een programma, nagaan of de doelen worden gerealiseerd, kunnen vele aspecten worden begrepen. Deze kunnen betrekking hebben op de organisatie, het budget, de specifieke resultaten op cliënt-niveau, enz. De verschillende (groepen van) belanghebbenden in de geestelijke gezondheidszorg hebben elk hun eigen doelstellingen: de cliënten zijn geïnteresseerd in de effectiviteit van hun eigen behandeling; zij streven vermindering van hun eigen klachten/problemen na. De medewerkers dienen resultaat te relateren aan hun inspanning; zij moeten zo efficiënt mogelijk hun aandacht verdelen over vele cliënt(system)en. De directie moet bewaken of met de beschikbare middelen de doelstelling van het instituut wordt bereikt; zij moet continuïteit waarborgen, zowel van het zorgsysteem als voor hun medewerkers. De overheid zal vooral zijn aandacht richten op het feit of aan de randvoorwaarden voldaan wordt, zowel financieel als organisatorisch. Binnen deze marges dienen dan zoveel mogelijk cliëntsystemen geholpen te worden.

In grote lijnen zullen de evaluatievragen zich richten op de *ingaaende* en *uitgaande* zijde van een programma, de "input" of "effort" (inspanning) en de "output" of "effect"-kant. Het tussenliggende gedeelte is het "process", de motieven voor het bereikte effect, zowel de meer oppervlakkige zoals motivatie van cliënten als de "*red light box*" van de behandelkamer van de therapeut.

Hiervan afgeleid maakt Suchman (1967) een onderscheid in vijf categorieën evaluatie-criteria volgens welk een programma geëvalueerd kan worden. Attkisson e.a. (1978) vervolmaken deze indeling. Deze vijf categorieën zijn:

1. *Inspanning ("effort")*. Dit betreft de kwaliteit en de kwantiteit van de "*input*" van het programma zoals aantallen cliënten, staf, budget, e.d. Ook de randvoorwaarden zoals wettelijke bepalingen waaraan voldaan moet worden, vallen hieronder. Alle informatie die aangeeft wat er in een programma gebeurt, wordt tot inspannings-evaluatie gerekend. Het zogenaamde profielpakket (Donker, 1984a) is een raamwerk voor deze evaluatierubriek. Voorbeeld van evaluatievragen zijn:

- Hoeveel (nieuwe) cliënten worden per tijdseenheid behandeld.
- Welke cliënten worden in behandeling genomen.
- Hoeveel geld kost de totale hulpverlening.
- Zijn de juiste medewerkers in dienst.

De belangrijkste informatiebron ter beantwoording van deze vragen zal een zgn. "*Management Information System*" kunnen zijn. Dit is "een systeem, dat gericht is op het systematisch verzamelen en verwerken van gegevens teneinde te voorzien in informatie, die relevant is voor het management van een organisatie" (Koopmans en Bouma, 1984, p. 183).

2. *Prestatie (Performance)* is de "output" van een programma. Het betreft hier de evaluatie van de resultaten, de effecten van de inspanning: welke veranderingen zijn er opgetreden en voldoen deze aan de voorafgestelde doelen. De bekendste evaluatievragen zijn:

- wat is de mate van verbetering van het functioneren van de cliënten;
- wat is het oordeel over de geleverde hulp;
- hoe effectief worden de verschillende beslissingen genomen (intake, ontslag).

Teneinde deze vragen te beantwoorden wordt vooral gebruik gemaakt van evaluatie vragenlijsten, die cliënten bijv. bij beëindiging van de hulpverlening invullen. Een goed voorbeeld om na te gaan of beoogde doelen worden bereikt is de Goal Attainment Scaling (Kiresuk & Sherman 1968, Melief e.a. 1979, zie ook hfst. 4.3.2).

Deze eerste twee categorieën bevatten enkelvoudige criteria: *input- of outputvariabelen*. De overige drie betreffen samengestelde criteria, relaties tussen input- of outputmaten met andere variabelen.

3. *Adequaatheid (Adequacy)* betreft de relatie tussen het effect en de (maatschappelijke) behoefte. Wanneer er een grote behoefte is aan een bepaalde voorziening, maar deze slechts voor weinig mensen toegankelijk is, wordt in totaal slechts een geringe maatschappelijke behoefte vervuld. Beoordeeld moet dan worden of een lagere effectiviteit niet meer mensen kan helpen en zo een betere verhouding effect/behoefte oplevert. Met druk van de overheid om bijv. de behandelingsduur in de geestelijke gezondheidszorg zo kort mogelijk te houden, lijkt een adequaatheidscriterium nagestreefd te worden. Wanneer in een bepaald programma per jaar 200 cliënten worden behandeld en dit is voor 50% effectief, levert dit 100 effectieve cliënten op. Wanneer bij een effectiviteit van 20% er in totaal 1000 cliënten behandeld kunnen worden, is het rendement 200 cliënten. Dergelijke rekenvoorbeelden zijn in zijn algemeenheid wel te maken, maar in de praktijk is de maatschappelijke behoefte niet meetbaar. Een groot bezwaar van een dergelijke benadering is ook, dat er wordt omgesprongen met rendementscijfers alsof het een lopende band is: de band 2x sneller laten lopen, levert een productieverhoging op van 100%; wordt een groter percentage misbaksels van 20% afgetrokken, dan is het verhoogde rendement toch nog 80%. In de geestelijke gezondheidszorg dient terdege rekening gehouden te worden met de consequenties van niet effectief behandelde patiënten. Het is eveneens een maatschappelijk belang om deze groep zo klein mogelijk te houden. Adequaatheid is derhalve een spanningsveld tussen effectiviteit en maatschappelijke behoefte en zal in onderhandelingen tussen bestuur en financier over cliëntenaantallen per jaar een belangrijke rol spelen.

4. *Doelmatigheid (efficiency)* betreft de relatie tussen inspanning en resultaat (input/output). Nagegaan wordt of met dezelfde inspanning niet méér bereikt kan worden of, in tijden van bezuiniging, met minder kosten hetzelfde resultaat behaald kan worden. Een vergelijkende studie van verschillende programma's kan informatie verschaffen over het meest efficiënte.
5. *Proces (process)-evaluatie* kan antwoord geven op de vragen *waarom* een bepaald programma deze resultaten bewerkstelligt. Volgens Suchman (1967) is dit niet een vast onderdeel van programma-evaluatie. *"An evaluation study may limit its data collection and analysis simply to determining whether or not a program is successful according to the preceding four criteria (effort, performance, adequacy and efficiency) without examining the why's and wherefor's of this success or failure"* (p. 66).

Hiermee zijn de vijf evaluatiegebieden gegeven. Het eerste gebied, het meten van de inspanning, is een eerste vereiste in een evaluatie-project, het tweede is noodzakelijk om de volgende drie criteria te kunnen operationaliseren. Een evaluatie-project dat volstaat met het meten van de inspanning en het resultaat, verschaft de primaire informatie om zowel intern als extern beleidsbeslissingen te kunnen nemen, hetzij om het programma te continueren of om het te wijzigen, terwijl er dan tevens indicaties zullen zijn in welke richting deze gezocht moeten worden.

Een andere indeling van typen van evaluatie geeft Kiresuk en Lund (1977). Zij delen evaluatie in vier gebieden in: *"Effort"*, *"Outcome"*, *"Efficiency"* en *"Impact"*, hetgeen min of meer overeenkomt met de eerste vier gebieden van Suchman. Hun onderverdelingen zijn echter zeer informatief, met name de verschillende vormen van *"outcome"*-evaluatie. Deze wordt gesplitst in twee rubrieken: gericht op het programma en gericht op de cliënt. Een dergelijk onderscheid is zeer relevant, omdat beide ook hun eigen doelstellingen hebben. Enerzijds zullen deze in elkaars verlengde liggen - het gezamenlijke c.q. gemiddelde doel van de cliënten is ook een van de doelen van het programma -, maar anderzijds kan het programma nog andere doelen nastreven welke niet gerelateerd zijn aan die van de cliënten.

Kiresuk en Lund onderscheiden drie modellen om een programma te evalueren:

- *"Goal model and goal free evaluation"*. Het *"goal model"* is het meest gebruikte model en wordt in par. 3.4 uitgebreid beschreven. Het *"goal free model"* is een vorm van externe evaluatie (zie 3.3.2), waarbij globaal de vraag moet worden beantwoord: *"Is the program any good"* (Kiresuk, 1977, p. 291). De externe evaluator laat zich hier niet door de doelen van de instelling zelf maar door het maatschappelijke belang leiden. De term *"goal free"* is in zoverre misleidend, dat dit belang eveneens vertaald kan worden in doelen welke moeten worden geoperationaliseerd. Het onderscheid interne-externe evaluatie is daarom duidelijker.

- "System model" is een vorm van procesonderzoek. Naast of na het evalueren van de doelen zal nagegaan worden waarom de doelen wel of (vooral) niet bereikt zijn.

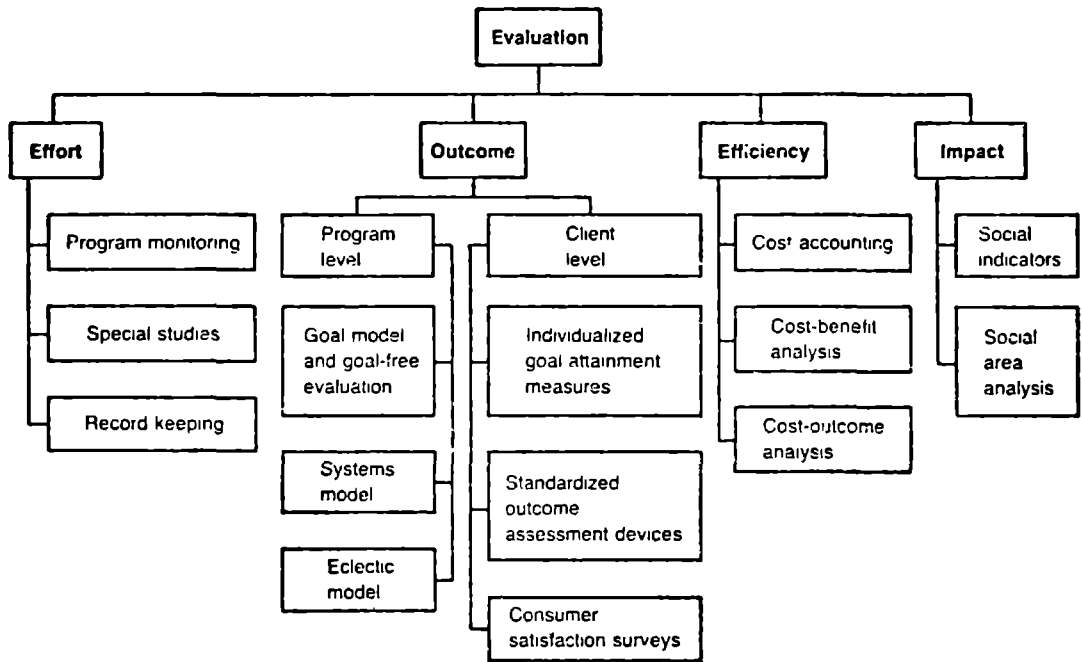
- "Eclectic model" is een hybride vorm van de drie bovenstaande evaluatiemodellen.

Op cliëntniveau worden drie meetprocedures genoemd:

- individuele doelmetingen zoals de G.A.S.;
- gestandaardiseerde uitkomstmetingen zoals psychologische vragenlijsten als M.M.P.I., F.S.S. e.d.;
- cliënt satisfactie metingen.

Deze metingen hebben elk hun eigen operationalisatiegebied en kunnen naast elkaar gebruikt worden.

In fig. 3.1 wordt de indeling van Kiresuk en Lund weergegeven. De overige gebruikte begrippen spreken voor zich.



Afb. 3.1 Indeling van typen van programma-evaluatie
(Kiresuk & Lund, 1977, p.291)

Samenvattend geeft Attkisson e.a. (1978) een overzichtelijke tabel van de diverse aandachtspunten waar programma-evaluatie zich op kan richten:

Programma-Evaluatie, een werkdefinitie:

Programma-Evaluatie is een proces om redelijke beoordelingen te maken van een programma ten aanzien van:

- inspanning
- effectiviteit
- efficiëntie
- adequaatheid.

Gebaseerd op een systematische gegevensverzameling en analyse. Opgezet om gebruikt te worden voor:

- programma management
- externe verantwoording
- toekomstplanning.

Met speciale aandacht voor:

- toegankelijkheid
- aanvaardbaarheid
- volledigheid
- geïntegreerdheid van diensten
- bewustzijn
- beschikbaarheid
- continuïteit
- kosten van diensten.

Hiermee zijn de inhoudelijke compartimenten van de container programma-evaluatie gevuld.

3.3.2 Organisatievormen van programma-evaluatie.

Tijdens een conferentie over programma-evaluatie in de geestelijke gezondheidszorg (Nijmegen, 1984) werd een serieus bedoelde maar door de deelnemers als mislukte scetch opgevatte discussie gehouden over twee zgn. diametraal tegenover elkaar staande vormen van programma-evaluatie: *interne* en *externe evaluatie*.

De achtergrond van deze tegenstelling is dat programma-evaluatie kennis oplevert welke ten goede of ten kwade kan worden gebruikt. Vooral deze laatste mogelijkheid wordt natuurlijk gevreesd en het is daarom van belang wie over deze kennis beschikt en wie bijgevolg de macht heeft om deze te gebruiken (Donker, 1984b).

De tweedeling interne versus externe evaluatie heeft betrekking op de vraag wie de evaluatie uitvoert en wie over de gegevens kan beschikken; blijven deze binnenskamers, in de eigen instelling (intern) of wordt het project uitgevoerd door instanties buiten de instelling zoals de overheid, financiers, e.d. (extern)? In de discussie tussen de conferentiegangers werd dit onderscheid als een *schijntegenstelling* gezien, immers de belangen van de verschillende partijen zijn

niet tegenstrijdig maar liggen in elkaars verlengde. Idealiter is dit ook zo, maar wanneer ongunstige resultaten worden aangetoond, zal dit druk van de externe instantie betekenen om het functioneren te veranderen en dit zal bij de hulpverleners op weerstand kunnen stuiten, hun positie, inhoudelijk of arbeidsrechtelijk, is dan in het geding.

Het onderscheid tussen interne en externe evaluatie kan op verschillende niveau's worden aangebracht:

- wie voert de evaluatie uit: een eigen medewerker of een buitenstaander;
- aan wie wordt gerapporteerd;
- wie stelt de evaluatiecriteria vast;
- wie kan de veranderingen bepalen.

Naarmate op meer niveau's externe instanties de verantwoordelijkheid hebben, zal de macht van de interne organisatie worden verminderd. Sommigen bepleiten dit (Beenackers, 1984a). Anderen menen, dat de eigen organisatie eerst zelf moet kunnen aantonen een goed beleid te kunnen voeren en ook impopulaire maatregelen durft te nemen.

- Aarsen (1984) noemt als voordelen voor interne evaluatie:
- het kan beter worden afgestemd op het eigen karakter van de instelling, de cliëntenpopulatie, e.d. Het kan daarom genuanceerder te werk gaan;
 - de resultaten zullen betrouwbaarder zijn, omdat de evaluator er zelf bovenop zit en er minder huiver bestaat om eerlijke antwoorden te geven die "eventueel tegen hen gebruikt kunnen worden";
 - het biedt betere garanties voor implementatie van de onderzoeksresultaten, de evaluator is met de organisatie vertrouwd en zijn resultaten zullen eerder geaccepteerd worden.

Hij concludeert dat externe instanties meer gericht zijn op productiviteit, terwijl de eigen organisatie vooral kwaliteit wenst te bewaken en deze is moeilijk extern af te dwingen.

Mastboom (1984) voegt nog een derde evaluatie vorm toe: *onafhankelijke programma-evaluatie*. Deze is niet gebonden aan een afzonderlijke instelling, maar regardeert het functioneren van de geestelijke gezondheidszorg als geheel: in hoeverre voldoet het zorgsysteem aan de totale maatschappelijke behoefte. De totale (landelijke of regionale) geestelijke gezondheidszorg wordt geëvalueerd. Hier kunnen dezelfde criteria gelden als voor interne of externe evaluatie, zoals bereikbaarheid, samenhang van diensten, e.d.

Schnabel (1984b) hanteert eveneens een driedeling, welke gebaseerd is op de reikwijdte van de evaluatie: een individuele behandeling, een bepaalde voorziening/verstrekking of een instituut/organisatie in zijn geheel. De evaluatie van een voorziening en van een instituut kunnen samenvallen wanneer een instituut slechts één voorziening heeft, bijv. alleen klinische psychotherapeutische hulpverlening.

Deze indelingsvormen overziend, blijkt er enige frixie te bestaan: onafhankelijke evaluatie is het hoogste aggregatieniveau op de dimensie: individueel - programmatisch -

institutioneel - onafhankelijk, en verdient eerder daar geplaatst te worden dan naast het onderscheid intern-extern. Worden deze twee indelingsprincipes gekoppeld, dan ontstaat het schema in figuur 3.2. Met een kruis zijn de combinatie-mogelijkheden aangegeven.

	individueel	programma	instituut	onafhankelijk
intern	x	x	x	-
extern	-	x	x	x

Fig. 3.2 Indeling programma-evaluatie-vormen

Wordt getracht programma-evaluatie-projecten in dit schema te plaatsen, dan is voor zover bekend het project Evaluatie Crisiscentrum Utrecht (Beenackers en Donker, 1981) het enige externe project op programma/instituutsniveau in Nederland. Tal van projecten zijn intern (Lange - Huize Padua, Walburg - Johannes Wier Huis, Hesselink - De Viersprong). Het programma-evaluatie-project dat in deel B wordt gepresenteerd, is eveneens een vorm van interne evaluatie op zowel individu- als programma/instituutsniveau, aangezien het instituut slechts één voorziening heeft. Ondanks de kritiek van Beenackers (1984a) op interne programma-evaluatie, is deze in Nederland de meest gebezigde evaluatie-methode. Of deze interne projecten de kwaliteit van de zorg in de geëvalueerde instelling zullen verbeteren of slechts dienen voor *public relations*, zoals Beenackers stelt, zal moeten blijken, alhoewel - wanneer ook de evaluatie van het project intern blijft - zal ook dat niet zichtbaar worden.

Opmerkelijk is overigens dat de bescheiden Nederlandse publicaties - voornamelijk het eerdergenoemde congresverslag - zich in grote mate bezighouden met de *organisatie* van programma-evaluatie en minder met de *inhoud* ervan, dit in tegenstelling tot de Amerikaanse publicaties, die bijna uitsluitend de methodiek van programma-evaluatie betreffen: het opzetten en uitvoeren van een programma-evaluatie-project en met name de meetbaarheid van de evaluatiecriteria en de benutting van de resultaten. Deze zijn zelfs uitgewerkt in volledige cursussen, zoals "Program Evaluation Kit" door Morris (1978, met in 1982 reeds de negende druk) en "Evaluation Primer" door Fink en Kosecoff (1978).

Dit gebrek aan voorbeelden in de Nederlandse geestelijke gezondheidszorg en onbekendheid met deze Amerikaanse stand van zaken, doen kennelijk Van Ham (1984) concluderen, dat programma-evaluatie geen oplossing biedt voor het doelstelings- en meetprobleem in de geestelijke gezondheidszorg. Dit zou zo zijn, wanneer deze zich alléén met de inspannings-

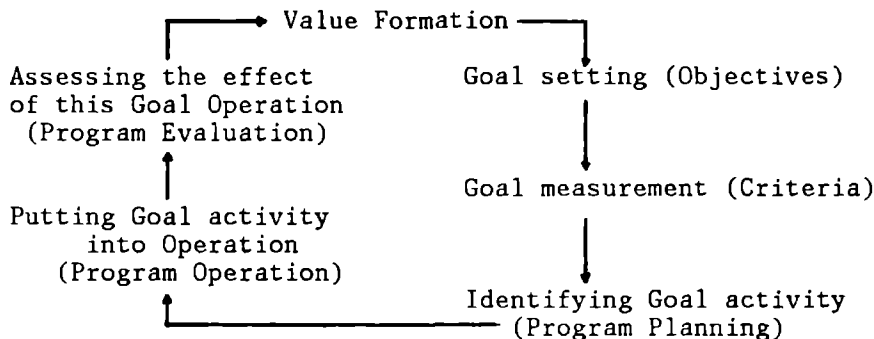
of "input"aspecten zou bezighouden. Dit zou dan misschien -zoals hij stelt- "meer van hetzelfde zijn" (p. 181), hoewel ook dat betwijfeld mag worden. Een doelgerichte evaluatie met systematische feedback aan de direct betrokkenen: beleidsfunctionarissen en/of managers, over de werkelijke effecten (de kwaliteit) van de behandeling, is meer dan in veel instellingen voor geestelijke gezondheidszorg nu gebruik is. Zijn (retorische) vraag "Is programma-evaluatie iets nieuws of gewoon goed management", laat zich als volgt beantwoorden: programma-evaluatie is zeker wat zijn doelstellingsgerichte evaluatie van de effecten van de behandeling betreft, iets nieuws en iedere goede manager zou zich gelukkig mogen prijzen van de resultaten van een dergelijke evaluatie gebruik te kunnen maken. Programma-evaluatie is meer dan een managementprocedure (cf. Management van non-profit organisaties van Ruiter, 1979), zoals uit deze paragraaf zal zijn gebleken. Overigens, is "*goed management*" (*in de geestelijke gezondheidszorg*) zo "*gewoon*"? Het is echter wel een voorwaarde om benutting van de feedback mogelijk te maken.

3.4 HOOFDLIJNEN VAN EEN PROGRAMMA-EVALUATIE-PROJECT.

Zoals ieder onderzoeks- of evaluatieproject, verloopt programma-evaluatie in een aantal stappen of fasen. Suchman (1967) beschreef er zes, waaruit bleek dat hem een dubbele vraagstelling voorstond: naast *evaluatie van doelstellingen* ook nagaan of de resultaten door het programma worden *veroorzaakt* of niet. Verder in zijn boek beschrijft hij dan ook (quasi) experimentele designs om dit te kunnen aantonen. In par. 3.2 is gesteld, dat hier als belangrijkste doel van een programma-evaluatie-project wordt gezien: het aantonen van de effecten van een programma in relatie tot de doelstellingen. Verantwoorde en generaliseerbare verklaringen voor het al dan niet bereiken van deze doelstellingen zijn op zijn best een positief neveneffect. Bij het opzetten van een programma-evaluatie-project kan met deze neveneffecten rekening gehouden worden, maar ze moeten beschouwd blijven als een bijproduct, dat overigens impliciet in het opgebouwde gegevensbestand aanwezig kan zijn. Het behandelingseffect van een geheel programma kan uitgesplitst worden naar effecten van onderdelen c.q. voor subgroepen van cliënten. Dergelijke analyses kunnen mogelijke verklaringen geven voor gevonden verschijnselen of minstens aanwijzingen in welke richting deze gezocht kunnen worden. In verder onderzoek kan dit dan worden nagegaan.

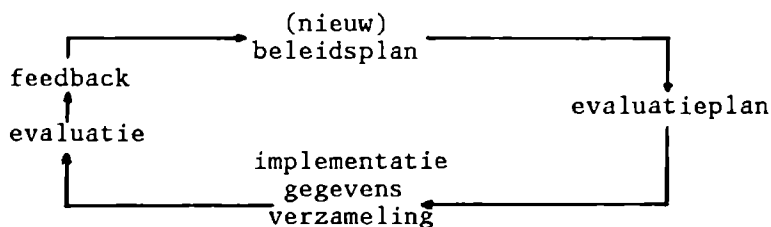
Een programma-evaluatie-project is een *circulair proces*. Suchman (1967) beschrijft dit proces in *doelactiviteiten* (zie fig. 3.3a): de eerste stap is het bepalen van een algemene waarde die aan het handelen ten grondslag ligt, bijv. gelukkig worden, lang leven, gezond zijn. Dit wordt vertaald in concrete doelen welke geoperationaliseerd worden in metingen. Het programma dient deze doelen te realiseren. Of dit lukt

wordt geëvalueerd en tenslotte wordt nagegaan of hiermee aan de oorspronkelijke algemene waarde is voldaan.



Afb. 3.3a Evaluation process (Suchman, 1967. p.34)

Wordt dit proces vertaald in een bestuurskundig of cybernetisch model, dan ontstaan als fasen (zie fig. 3.3b): *beleidsplan*, *evaluatieplan*, *implementatie*, *evaluatie* en *feedback*. Vervolgens kan deze leiden tot een nieuw beleidsplan, waarna de cyclus zijn volgende gang doorloopt. In tegenstelling tot de opvatting van Van Ham (1984) stopt deze cyclus niet als wordt aangetoond dat beleid en praktijk overeenstemmen. Het is een bewakingssysteem, en dat stopt niet als er niets (meer) aan de hand is. Programma-evaluatie is een continu proces, dat een integraal onderdeel is of dient te zijn van de organisatie (Attkisson & Broskowski, 1978, p. 22).



Afb. 3.3b Programma-evaluatie-proces

In het kort worden hier de belangrijkste fasen beschreven.

3.4.1 Het beleidsplan.

Als eerste moet geëxpliciteerd worden wat de *doelen* van het programma zijn. Deze doelen zijn vaak impliciet aanwezig en kunnen op verschillende niveau's gelegen zijn. Als het programma het (curatieve) functioneren van een heel instituut betreft, kunnen hierin afzonderlijke groepen van belanghebbenden onderscheiden worden met elk hun eigen doelen en bijgevolg behoefte aan informatie en evaluatie. Het beleidsplan wordt omgezet in een concrete c.q. objectieve informatiebehoefte per niveau of groep die deze informatie zal gebruiken ter evaluatie van dit plan.

Op *cliënt- en therapeutniveau* zal de vraag gesteld kunnen worden naar de vooruitgang en effectiviteit van de behandeling: bereikt de individuele cliënt de doelen waarvoor deze in behandeling is gekomen, wordt tijdig voortgezette hulp gerealiseerd, wordt aan de gemaakte afspraken voldaan, etc.

Op *teamniveau* kunnen deze individuele gegevens geaggregeerd worden, waardoor het (gemiddelde) teamfunctioneren geëvalueerd kan worden. Ook algemenere doelen welke het team zich gesteld heeft, zoals de aard en de duur van de behandeling, de doelgroep, e.d. dienen in het beleidsplan van het team geformuleerd te worden.

De *directie* vervolgens zal enerzijds geïnteresseerd zijn in de gesommeerde behandelingsresultaten, anderzijds zal zij informatie wensen over het bereiken van de beoogde cliëntenpopulatie, de efficiëntie van de organisatie, de satisfactie van de cliënten, de juistheid van de beslissingsprocedures zoals intake en ontslag, en dergelijke.

Informatie over algemene instituutsdoelen zoals aantal en aard van de behandelde cliënten, behandelingsduur en -effect zal van belang zijn voor *externe instanties*. Zij wensen de doelen te evalueren die zij beogen met het financieren van het instituut. Interne en externe doelen liggen in elkaars verlengde. Een verschil kan ontstaan bij het bepalen van het preciese criterium. De behandelingsduur bijv. is zo'n twistappel. De financier wenst deze zo kort mogelijk te houden, terwijl de behandelaars omwille van een zo effectief mogelijke hulpverlening, deze juist wensen te verlengen. Ten aanzien van dergelijke verschillen is het van belang onderscheid te maken tussen interne of externe programma-evaluatie. Maar beide partijen zullen gezamenlijk of afzonderlijk de criteria moeten bepalen volgens welke deze moeten worden geëvalueerd. Indien externe instanties eveneens van evaluatieve informatie moeten worden voorzien, kan een spanningsveld ontstaan tussen de interne en externe belanghebbenden. De positie van de evaluator en de doelstellingen van het programma-evaluatieproject zijn in deze van cruciaal belang; als er sprake is van twee tegenstrijdige heren moet vooraf duidelijk zijn welke heer moet worden gediend. Op ieder niveau moeten de doelen die moeten worden geëvalueerd, *expliciet* en *transparant* zijn voor alle betrokkenen, alvorens de volgende fase kan starten.

3.4.2 Het evaluatieplan.

Naarmate de doelstellingen en criteria in het beleidsplan preciezer geformuleerd zijn, is het evaluatieplan des te eenvoudiger. Volgens Rutman (1980) is *"a well defined program a precondition for an effectiveness evaluation"* (p. 44). In dit plan wordt vastgelegd welke informatie door wie op welk moment zal worden verzameld en op welke wijze deze gegevens zullen worden verwerkt. Hierbij moet nauw worden aangesloten bij de bestaande situatie en procedures. De evaluator kan de praktijksituatie niet naar zijn hand zetten. Hij moet een meet- en regelnetwerk aanleggen door een bestaande organisatie, zoals het aanleggen van een nieuwe pijpleiding door een reeds bestaand netwerk. Sommige leidingen kunnen en mogen afgebroken worden of kunnen vervallen omdat de nieuwe leiding hun taak overneemt. Andere mogen echter niet worden aangevoerd (Sijben, 1980b, p. 12).

Afhankelijk van de veelheid en diversiteit van doelstellingen kan de te verzamelen informatie zeer uitgebreid zijn. Zeker wanneer naast inspanningsvariabelen ook prestatiegegevens moeten worden verzameld, hetgeen doorgaans het geval zal zijn. Het zal veelal niet mogelijk of wenselijk zijn om alle meetinstrumenten tegelijkertijd in te voeren. Een geleidelijke opbouw van het evaluatieplan is verkieslijker, waarbij eerst begonnen wordt met het inwinnen van informatie die weinig tijdrovend en bedreigend is, zoals systematische registratie van patiëntengegevens, en die bruikbaar is, bijv. voor jaarverslagen, e.d., en doelen op dat niveau evalueren zoals behandelingsduur, doelgroep, enz.

De volgende stap kan zijn het invoeren van effectmetingen en/of registratie van interne procedures (intake, ontslag, nazorg regelen). Geleidelijk worden zo over alle instituutsdoelen middels meetinstrumenten gegevens verzameld, *van globaal naar specifiek, van oppervlakkig naar indringend, van registratie naar evaluatie van de behandeling*. Deze stappen worden in het evaluatieplan beschreven. De meetinstrumenten zelf volgen uit de geëxpliciteerde doelen. Naarmate deze concreter zijn, is het eenvoudiger deze te operationaliseren. Dit geldt met name voor de effectdoelen, het meest moeilijke onderdeel van het evaluatieplan. Het formuleren hiervan is enerzijds een inhoudelijke vraag die behandelaar en cliënt dienen te expliciteren, anderzijds een methodologisch probleem: hoe worden deze effecten het meest betrouwbaar en valide geoperationaliseerd. Dit laatste komt in hoofdstuk 4 uitgebreid aan de orde.

3.4.3 De implementatie.

Voor de geleidelijke invoering van het evaluatieplan wordt het zgn. *ui-model* voorgesteld (Sijben, 1977a, zie ook hoofdstuk 6). Het functioneren van een organisatie wordt schilsgewijze geëvalueerd: de buitenste schillen zijn het meest

eenvoudig te pellen, ze zijn minder interessant (sappig), maar wel noodzakelijk om bij het binnenste van de ui te komen. In de implementatiefase worden één voor één deze schillen gepeld: de meetprocedures worden geleidelijk ingevoerd. Hierbij kunnen zich uiteenlopende problemen voordoen: In de praktijk kan blijken, dat bepaalde procedures te arbeidsintensief zijn, inhoudelijk op teveel weerstand stuiten of dat ze niet die informatie opleveren die bedoeld was. Het overwinnen van weerstanden, het veranderen van het instrumentarium, het informeren en begeleiden van de gegevensverstrekkers of -verzamelaars is het belangrijkste werk van de evaluator in deze fase. Hiervan zal voor een groot gedeelte het welslagen van het evaluatieproject afhankelijk zijn.

3.4.4 De evaluatie en feedback.

De in de vorige fase verzamelde gegevens dienen periodiek geanalyseerd en gerapporteerd te worden, overeenkomstig de specificaties in het evaluatieplan. De analyses kunnen zeer uiteenlopend zijn. Soms kan volstaan worden met eenvoudige beschrijvende parameters, zoals percentages en gemiddelden om het al dan niet realiseren van de doelstellingen aan te tonen. Doorgaans is middels bivariate of multivariate analysetechnieken een nadere aanduiding te geven van verbanden welke in de gegevens besloten liggen, zoals verschillen tussen subgroepen (diagnose, behandeling, geslacht, e.d.). Dergelijke analyses zijn eerder *exploratief* dan *verklarend*, maar kunnen behulpzaam zijn bij het zoeken naar causale relaties. Exploratieve analysetechnieken onderscheiden zich van de toetsende methodieken door niet uit te gaan van een bepaald model waaraan de gegevens zouden moeten voldoen, maar "slechts" patronen in de gegevens op te sporen. "The model is not brought to the data, it is extracted from it" (McGaw, 1981, p. 73)

Na deze analyses volgt de *rapportage*, de feitelijke feedback aan de opdrachtgever. Dit sluitstuk van het evaluatieproject bevat de evaluatie van de doelstellingen. Hoewel het niet tot de verantwoordelijkheid van de evaluator kan worden gerekend, dat deze resultaten ook worden gebruikt, zal deze wel ervoor dienen te zorgen dat ze gebruikt kunnen worden. Programma-evaluatie wordt ook aangeduid met "*utilization-focused evaluation*" (Patton, 1978) of "*Utilizing evaluation*" (Ciarlo, 1981). In het laatstgenoemde boek worden onderzoeksprocedures beschreven om na te gaan of de evaluatiegegevens gebruikt zijn bij nieuwe beleidsbeslissingen. Er wordt zelfs een *benuttingsschaal* gepresenteerd. Bepleit wordt, dat benuttingsstudies een onderdeel moeten zijn van programma-evaluatie. De taak van de evaluator is dan met de periodieke rapportage niet ten einde. Naast de wijze van de presentatie van de gegevens zal het feit dat een dergelijke studie volgt de benutting ervan doen toenemen. Opmerkelijk is, dat derge-

lijke studies nodig zijn: programma-evaluatie is kennelijk een activiteit die weerstanden en verdringing oproept, terwijl deze toch tot doel heeft het functioneren te verbeteren. In de paragraaf 3.5 wordt op deze aspecten van programma-evaluatie nader ingegaan.

3.4.5 Een nieuw beleidsplan.

De cirkel van het evaluatieproces is nu gesloten. Indien de resultaten daar aanleiding toe geven en de organisatie of opdrachtgever bereid is deze te gebruiken, wordt het beleidsplan veranderd. Dit kan betekenen, dat het programma c.q. de activiteiten om de doelen beter te realiseren, worden gewijzigd. Het is ook mogelijk, dat de doelen veranderen, omdat deze niet haalbaar blijken te zijn. Dit wordt wel eens gezien als de zwakke kant van doelgerichte evaluatie: met *flexibele* doelen zou immers *altijd* effectiviteit kunnen worden aangetoond. Deze flexibiliteit is echter aan grenzen gebonden; de cliënten enerzijds en de financiers anderzijds zullen invloed hebben op bijv. de hoogte van het effectiviteitscriterium. Zelfs bij bijstelling van de doelen in plaats van het veranderen van het beleid, is de winst van het evaluatieproces, dat dit aan de belanghebbenden bekend wordt. Kennelijk werd voordien gemeend, dat hogere doelen werden gerealiseerd dan in werkelijkheid het geval was. Deze duidelijkheid is van belang voor de beoordeling van de adequaatheid van het totale programma.

3.5 EVALUEREN EN GEEVALUEERD WORDEN.

In de vorige paragrafen werd reeds melding gemaakt van de *interactie* tussen evaluator en geëvalueerde. De praktijkgerichtheid van programma-evaluatie maakt nauw contact en overleg tussen de onderzoeker en de onderzochte instelling noodzakelijk in alle fasen van het evaluatieproces. Willen uiteindelijk de resultaten gebruikt (kunnen) worden, het hoofddoel van het evaluatieproject, dan zal aan meerdere voorwaarden voldaan moeten zijn: onderzoekstechnische en interactionele. De onderzoekstechnische aspecten zijn in het voorafgaande beschreven, op de interactionele wordt hieronder nader ingegaan.

Beenackers (1984b) geeft een opsomming van problemen welke kunnen optreden tussen onderzoeker en onderzochten in de diverse fasen van het evaluatieproces. Hoewel deze vooral gebaseerd zijn op ervaringen met een extern evaluatieproject zijn deze, eventueel in mindere mate, ook van toepassing op interne projecten. De problemen worden door hem benoemd als "*weerstand*" van de onderzochten tegen de onderzoeker en zijn werkzaamheden. Deze weerstanden betreffen: het beoordeeld worden, de vraagstelling, de evaluatieprocedure, de uitkoms-

ten en het gebruikmaken van deze uitkomsten intern (beleid) en extern (publicatie). Uiteraard zullen deze weerstanden groeien, wanneer er sprake is van belangentegenstellingen, zoals dit bij externe evaluatie meer het geval is. Patton (1978, 1980, 1981) staat meer een interne vorm van evaluatie voor, waarbij de belangrijkste evaluatiebeslissingen genomen worden door een team van medewerkers en de evaluator. Door deze spreiding van verantwoordelijkheden wordt vermeden, dat de weerstand zich richt op de evaluator, die door zijn positie reeds een buitenstaander is. Blöte e.a. (1983) beschrijven drie projecten waarbij volgens deze strategie gewerkt is, met positief resultaat. Het risico van een dergelijke evaluatie is echter, dat kritische gegevens weinig kans maken in het evaluatieplan opgenomen te worden. Zo zijn zij van mening, dat gevoelige gegevens zoals personeelsgebonden informatie vermeden moeten worden, of pas na goede afspraken gepresenteerd kunnen worden. Hoe dergelijke informatie is benut in het door hen beschreven project, wordt niet vermeld.

Deze twee benaderingen staan diametraal tegenover elkaar: de *democratische* evaluatievorm van Patton c.s. en de *hiërarchische* van Beenackers. Patton tracht de instelling zover te brengen, dat ze zichzelf evalueert, maar gaat voorbij aan het feit dat zelfkritiek niet een eigenschap is, die doorgaans sterk aanwezig is bij personen en instanties. Beenackers meent daarom dat deze kritiek van buitenaf gegeven moet worden. Oplossingen voor de opgeroepen weerstand zijn er eigenlijk niet, ze kunnen hoogstens iets worden verzacht. Benutting van de gegevens zal volgens hem dan ook alleen kunnen, wanneer dit wordt afgedwongen met sancties. De onderzoeker mag zelf geen concessies doen in de belangenstrijd met de onderzochten. Zijn positie moet zo "sterk gemaakt worden dat de onderzochten aan het kortste eindje trekken" (p. 141).

Met Beenackers wordt hier geconcludeerd, dat er nog te weinig ervaring is c.q. gepubliceerd is over de problemen tussen deze twee partijen, ze zijn er wel, maar er wordt - zeker openlijk - over gezwegen. "Het gevolg is dat er vele irrealistische verwachtingen in stand gehouden worden, en dat onderzoekers en onderzochten telkens opnieuw door schade en schande wijs moeten worden. En soms is dit zo frustrerend dat van verder onderzoek maar wordt afgezien" (Beenackers, 1984b, p. 142), luidt zijn sombere voorspelling aan het slot. Om dit te voorkomen zou bij ieder evaluatierapport een beschrijving van de *interactie(problemen)* moeten worden opgenomen. Of een dergelijke publicatie inderdaad gerealiseerd kan worden is echter de vraag, gezien het gevoelige karakter van deze informatie. Ongewenste publiciteit wordt zelden geduld (zie De Zwaan: "Met de prop in de mond - over de vrijheid van onderzoek" (1983)). Een oplossing voor de problemen van elk van de twee beschreven rolopvattingen van de evaluator is niet te geven. Ook de richtlijnen voor "program evaluation" van de Evaluation Research Society (de Brey, 1981) geven hierover geen enkele informatie. Als laatste richtlijn wordt slechts vermeld, dat de evaluator een duidelijk onderscheid

moet maken tussen zijn evaluatieve rol en iedere andere rol waarbij gekozen is voor de verdediging van een standpunt.

Weerstand tegen de resultaten c.q. tegen het gebruik ervan is niet voorbehouden aan hiërarchische c.q. externe programma-evaluatie. Ook interne evaluatoren kennen de weerstand tegen de evaluatie en de resultaten. Hesselink (1981) meldt vanuit zijn ervaring als interne onderzoeker drie mogelijke reacties op negatieve resultaten:

1. *Wegverklaren*: verklaren *waarom* de cijfers zo zijn, dan hoef je er niets meer aan te doen. Bijv.: er zijn meer mensen met een W.A.O.-uitkering opgenomen en die zijn nu eenmaal moeilijker te behandelen.
2. *Ontkennen*: het klopt niet, de cijfers zijn onjuist. Het onderzoek is niet goed (uitgevoerd). Een duidelijk voorbeeld hiervan deed zich voor in een programma-evaluatie project bij een RIAGG/CAD te Amsterdam. In het onderzoek werd aangetoond dat van de 155 aangemelde patiënten reeds 60 gedurende de intake-fase waren uitgevallen. Een gedeelte bleek administratief te zijn "verdwenen". Door de medewerkers werd toen "het vertrouwen in het onderzoek opgezegd" (mondelijke mededeling Romijn, Psychol. Lab. Nijmegen, zie ook Romijn, 1984). Een ander voorbeeld staat vermeld in hoofdstuk 11.
3. *Overvragen*: Er is méér onderzoek nodig, en in tien minuten worden meer onderzoeksvragen gesteld dan in een jaar zijn te beantwoorden. Met het trekken van conclusies en het doorvoeren van beleidsveranderingen wordt gewacht tot de onderzoeker (afdoende) verklaringen heeft gegeven.

Voor de derde reactie is reeds eerder gewaarschuwd en deze kan gedeeltelijk voorkomen worden door de taak van de evaluator duidelijk te omschrijven. De eerste vorm van weerstand gaat voorbij aan de gestelde doelen; deze moeten gerealiseerd worden ondanks bepaalde patiëntenparameters. Een verandering in het patiëntenbestand zou niet *zonder meer* mogen leiden tot een verminderde effectiviteit. Het behandelingsbeleid zou hieraan moeten worden aangepast. Een wellicht oneerbiedige vergelijking is die met een garage. Als de auto's complexer worden is dit geen excuus om minder effectief te repareren. Uiteindelijk zullen dan alleen nog banden kunnen worden opgepompt en lampen vervangen, omdat dit het enige is wat nog hetzelfde is gebleven. Andere auto's vereisen andere reparatietechnieken en bijgeschoold personeel. Het is de taak van de garagehouder om hiervoor te zorgen. Een verklaring is geenszins een oplossing of alibi.

De ontkennende vorm van weerstand is het moeilijkst te bestrijden. Discussie is dan niet meer mogelijk, onderzoeker en onderzochte belanden in een "wellis-nietis" patstelling en naar de resultaten zelf wordt niet meer gekeken. Programma-evaluatie is primair een *spiegel* die een instelling wordt voorgehouden. Het is aan de medewerkers om het gepresenteerde beeld te accepteren en de benodigde schoonheidsbehandeling te geven. Of dit onder dwang of uit vrije wil moet

gebeuren, is (nog) niet te zeggen. Publicatie van met name Nederlandse programma-evaluatie-projecten zal moeten leren welke manier het beste werkt. Het stukgooien van de spiegel is in ieder geval de slechtste oplossing, een overigens herkenbare regressieve reactie die niet mag worden uitgesloten.

Gelet op de ervaringen van programma-evaluatoren en bovenbeschreven reactiewijzen tegen de resultaten, kan de eerste van de 96 (!) stellingen waarmee Cronbach zijn boek "Toward Reform of Program Evaluation" (1980) begint, beter worden genuanceerd: "*Program evaluation is a proces by which society CAN learn about itself*". Of de maatschappij in het algemeen en (G.G.Z.) instellingen in het bijzonder *willen* leren, is een noodzakelijke voorwaarde. Of programma-evaluatoren goede leraren zijn c.q. een goed leerboek (kunnen) schrijven, eveneens.

Hoofdstuk 4:

Het meten van therapie-effecten

If the proverbial Martian were to review the therapy outcome literature, he would undoubtedly conclude that earthmen embark upon expensive and timeconsuming programs of treatment to effect modification in their Rorschach, T.A.T. or M.M.P.I.-responses...
Bandura (1969) p. 53

4.1 INLEIDING.

In een voordracht op de jaarvergadering van Amsterdamse M.O.B.'s in 1980 besprak ik de mogelijkheid tot meten van behandelingsresultaten door middel van doelevaluaties, met name de Goal Attainment Scaling. Na deze lezing stond een van de toehoorders verontwaardigd op en wierp de volgende opmerking de zaal in: "Meneer, met al dat meten probeert u het menselijk geluk plat te slaan tussen de bladeren van een boek. Wij proberen onze cliënten *gelukkig* te maken, en menselijk geluk is niet te meten." Met deze reactie wordt in essentie het probleem weergegeven van de *operationalisatie van therapie-effecten*, en het protest van therapeuten tegen al datgene wat onderzoekers als effectmeting menen te kunnen gebruiken en dat niet in overeenstemming is met wat zij met hun behandeling beogen te bereiken. Onderzoekers hebben de neiging dergelijke reacties als *weerstand* uit te leggen, en niet altijd ten onrechte.

Als bovenstaande mening exemplarisch zou zijn voor de totale groep hulpverleners in de geestelijke gezondheidszorg zou evaluatie van behandeling per definitie uitgesloten zijn. Inderdaad, menselijk geluk is niet te meten en onderzoekers reduceren de werkelijkheid tot een tweedimensionele afbeelding: de pagina's van hun publicaties. Op de conferentie: *Programma-evaluatie in de geestelijke gezondheidszorg* (zie hoofdstuk 3.3.2) werden de resultaten gepresenteerd van een enquête onder de deelnemers (Sijben, 1984b). Eén vraag luidde: *In de geestelijke gezondheidszorg is het werkelijke effect nooit aantoonbaar*". Van de 169 invullers was 32% het met deze uitspraak eens, 6% had geen mening en 62% was het hiermee oneens. De Amsterdamse reactie komt "slechts" overeen met de mening van een derde van de responderende conferentie-gangers. Dit is natuurlijk geen representatieve steekproef uit de populatie van hulpverleners.

Als er van wordt uitgegaan, dat effecten inderdaad *aantoonbaar* kunnen zijn - wat onderzoekers natuurlijk reeds a priori doen - is het probleem *hoe* deze dan geoperationali-

seerd moeten worden. Deze vraag leidt tot een volgende: Wat doen therapeuten, welke effecten streven ze na met hun behandeling? *"Vraag aan psychotherapeuten wat ze doen, en je krijgt een bonte verzameling antwoorden. Dat onder psychotherapie zoveel verschillende dingen worden verstaan, maakt evaluatie van de gang van zaken niet gemakkelijker"* schreef Sou-dijn (1976, p. 78), het evalueren van psychotherapieën tot een van de *dilemma's* in sociaal wetenschappelijk onderzoek bestempelend.

In dit hoofdstuk wordt op het operationaliseren van effecten van psychotherapie nader ingegaan. Dat dit hoofdstuk een plaats vindt in dit boek over programma-evaluatie vindt zijn verklaring in het feit dat het evalueren van behandelingsresultaten in de geestelijke gezondheidszorg er niet aan ont-komt zich in dit dilemma te begeven. *Valide effectmaten is een vereiste voor een goed evaluatieproject.*

4.2 HET OPERATIONALISATIEPROBLEEM.

Het methodologische probleem om valide uitkomstmaten te construeren heeft vele onderzoekers in de loop van de jaren bezig gehouden en heeft vele haken en ogen. De gekozen operationalisatie heeft directe gevolgen voor de validiteit van het onderzoek. Invalide resultaten leiden vervolgens tot een geringe invloed van het psychotherapie-onderzoek op de praktijk.

Malan en London (1973) concludeerden: *"Surely much of this state of affairs (the lack of impact of research on clinical practice) is due to a single factor, namely the failure to design outcome criteria that do justice to the complexity of the human personality"* (p. 728). Als deze conclusie letterlijk zou worden gevolgd, is evaluatie van psychotherapie welhaast uitgesloten; wanneer zal ooit recht gedaan worden aan de complexiteit van de menselijke persoonlijkheid? Of de evaluatiecriteria daaraan moeten voldoen, mag echter worden betwijfeld, doorgaans zullen de effecten niet zo indringend zijn dat ze de persoonlijkheid aantasten. Maar hier zal per psychotherapeut wel verschillend over geoordeeld worden, eventueel in relatie tot de psychotherapeutische richting welke beleden of gepraktiseerd wordt. In het licht van de uitspraak van Bandura aan het begin van dit hoofdstuk zouden dergelijke operationalisaties kunnen leiden tot indirecte, abstracte effectmetingen die evenmin zicht geven op de resultaten van de behandeling. Bandura (1969) bepleit effectmetingen die dicht bij het te behandelen probleem aansluiten. (p. 53 e.v.)

Het operationalisatie-probleem kan eerder aldus gedefinieerd worden: *wanneer kan er van een daadwerkelijk resultaat* gesproken worden? De discussie over statistische significantie en klinische relevantie bestaat reeds ruim 30 jaar (Meehl, 1954). Onderzoekers gebruiken de statistiek als maatstaf voor het effect, de klinici hun oordeel. Dijkhuis

(1960) probeert deze te verenigen: "*Beschikken wij over een gestandaardiseerd en geformaliseerd criterium, dan is het waarschijnlijk beter om objectieve tests te gebruiken; is het criterium niet gestandaardiseerd en alleen uitlegbaar in kwalitatieve descripties, dan zullen naast kwantitatieve testgegevens ook 'clinische' beoordelingen een belangrijke rol kunnen vervullen*" (p. 23). Aangezien het criterium bij psychotherapie niet gestandaardiseerd is, zou dus voor deze dualistische benadering gekozen moeten worden. Deze aanbeveling is zelden opgevolgd.

Onderzoek dat primair bedoeld is om de resultaten van een behandeling vast te leggen, zoals programma-evaluatie, hoeft zich niet zozeer te richten op de *verklaring* voor de opgetreden veranderingen; het dient deze "*alleen maar*" *aan te tonen*. Om dit te kunnen doen, moet vooraf bepaald worden welke verandering wordt beoogd. De psychotherapeut zal deze doelen *vertalen* volgens zijn therapeutisch kader in onderliggende motieven, aanleidingen, oorsprong van de klachten, e.d. Deze theoretische constructie leidt tot een behandelingsplan. Aan het slot dient dit model aan de praktijk te worden *getoetst*. Als de problemen veroorzaakt werden of verklaard konden worden vanuit deze theoretische constructie, dan zal dit moeten blijken uit een verbetering van de problemen waarmee de patiënt met de behandeling is begonnen. Blijken deze aan het slot *niet* verbeterd te zijn, dan is óf de vertaling onjuist geweest óf de daarop aansluitende behandeling niet goed uitgevoerd. Een reactie van de therapeut dat de behandeling wél succesvol is geweest, maar dat de patiënt dit niet kan zien vanwege zijn problematiek, is eveneens een bewijs dat de behandeling (grotendeels) ineffectief is geweest: de problematiek waarmee de patiënt startte, bestaat kennelijk nog steeds. Mintz (1981) stelt dan ook terecht: "*If a patient's initial symptoms do not improve, there will in general be a good consensus that treatment has not been successful*" (p. 506). Dit pleit voor het gebruik van effectmetingen die direct afgeleid zijn van de klachten of problemen van de patiënt.

De literatuur over psychotherapie-effect-onderzoek overziend, blijkt dat dit kennelijk geen voor de hand liggende conclusie is. Hoewel het aantal gebruikte meetinstrumenten zeer talrijk is, komen dergelijke operationalisaties weinig voor, zoals blijkt uit de meta-analytische studies welke de laatste jaren in zwang zijn geraakt. Deze onderzoeken recapitulerend, somt Schagen (1983, p. 80 e.v.) diverse indeelingsprincipes op van therapie-effect-metingen. Deze zijn gebaseerd op:

- de *inhoud* van de meting zoals angst, depressie, lichamelijke klachten;
- de *reactiviteit* of beïnvloedbaarheid door degene die de meting verricht of gemeten wordt;
- de *specificiteit* van de meting, i.c. de relatie tussen de maat en datgene wat gemeten wordt. Doelmetingen hebben een hoge specificiteit, projectieve tests een lage.

Strupp en Bergin (1969, p. 55) bepleiten het gebruik van meerdere instrumenten naast elkaar. Bergin en Lambert (1978, p. 171 e.v.) geven een aantal aanbevelingen en suggesties voor de keuze van therapie-effect-metingen. Deze dienen volgens hen de potentiële verandering *multidimensioneel* te meten. Door therapie ontstaan *divergente* veranderingsprocessen waar iedere patiënt op zijn eigen manier op reageert. Door *divergente* metingen zou de kans worden vergroot, dat het effect kan worden aangetoond. Een probleem hierbij ontstaat, wanneer deze variëteit aan instrumenten verschillende resultaten geven: de ene toont wel effect aan en de andere niet. Welke conclusie moet dan getrokken worden, of telt altijd die meting, die het meeste effect weergeeft? Een meting die geen effect laat zien, kan immers ook een juist beeld geven. Bergin en Lambert stellen ook voor *diverse informanten* te gebruiken: patiënt, behandelaar, externe beoordelaar en significante ander. Hun voorkeur gaat uit naar gestandaardiseerde meetinstrumenten, zoals M.M.P.I., P.O.I. (Shostrom, 1971). Zogenaamde "Home made" vragenlijsten om bij beëindiging van de behandeling het effect ervan te meten, verschaften volgens hen weinig extra informatie vergeleken met bijv. de M.M.P.I. Opmerkelijk is, dat zij *doelevaluatie-methoden* door de cliënt kennelijk niet als een mogelijke effectmeting zien. Voor de therapeut wordt de zogenaamde "*target-complaints*"-methode (Battle e.a. 1966, Sloane e.a. 1975) wel als evaluatie-instrument genoemd.

In de meta-analytische studies van Smith & Glass (1977) en Shapiro & Shapiro (1982, 1983) worden effectmetingen die gebaseerd zijn op door de patiënt geformuleerde doelen, evenmin genoemd. Al de geanalyseerde onderzoeken betreffen experimentele, vergelijkende therapie-effect studies. De slotconclusie dat "*contemporary outcome research is not representative of clinical practice*" (Shapiro & Shapiro, 1982, p. 598) spreekt voor zich. In tien jaar is sinds Malan (1973) in dat opzicht kennelijk nog niets veranderd. In hoofdstuk 3 is hiervoor een verklaring gegeven. Een andere onderzoeksmethodiek: van *demonstratie* van effecten naar *registratie* hiervan (Veerman, 1983) vraagt ook om andere evaluatie-metingen. Onderzoek dat nauw moet aansluiten op de praktijk moet hierin ook de patiënten betrekken. Zij zijn de eerst aangewezenen om het resultaat van hun behandeling te beoordelen. Reeds in 1970 betoogt Fiske e.a. dat teveel nadruk gelegd wordt op het gebruik van uitkomstmaten die voor alle patiënten gelijk zijn. "*The patient typically comes in with one major complaint for which he seeks relief. Similarly, the therapist may develop particular outcome goals for each patient. Should not outcome be defined to be congruent with such individual aims? The methodology for studies with this orientation needs further development*" (Fiske e.a., 1970, p. 24). Een suggestie die binnen het psychotherapie onderzoek weinig is opgevolgd, afgezien van bijv. het onderzoek van Sloane (1975). Uiteraard kunnen de oordelen van patiënten vergeleken worden met die van hun behandelaars, maar diverse auteurs geven aan dat deze laatsten niet de meest betrouwbare

informatiebron zijn. Smith & Glass (1977, p. 755) kennen aan deze metingen de hoogste *effectiviteitsscore* toe. De waarde van het klinische oordeel wordt hierdoor sterk gedevalueerd.

Het voorgaande in overweging nemende lijken een drietal evaluatie-methoden geschikt, of worden althans door diverse auteurs aanbevolen. In een registratie van behandelingsresultaten zouden deze naast elkaar moeten worden gebruikt en onderling worden vergeleken:

- Doelevaluatie methoden.
- Psychologische tests/vragenlijsten die aansluiten bij de te behandelen klachten of problemen.
- Subjectieve beoordelingen van het behandelingsresultaat en de tevredenheid hierover.

Per methode wordt in paragraaf 4.3 een beschrijving gegeven van de belangrijkste kenmerken.

Naast de *aard* van de operationalisering is ook de *mate* van de beoogde of geconstateerde verandering een probleem. Genoemd is al het onderscheid tussen een klinisch relevant of een statistisch significant verschil. Dit *criteriumprobleem* is zonet nog moeilijker op te lossen, wanneer geen gebruik gemaakt wordt van de klassieke t-toets om verschillen tussen voor- en nameting te evalueren. Deze toets lijkt welhaast de "gouden standaard" ter evaluatie van het effect. Wordt deze niet gevolgd, dan ligt een vergelijken met normgroepen voor de hand. Zo worden in Denver middels een speciale vragenlijst (Denver Community Mental Health Questionnaire) de resultaten van de patiënten na 90 dagen follow-up vergeleken met een steekproef uit de normale populatie (zie Ciarlo en Davis, 1976; Ciarlo & Reihman, 1978; Butters e.a., 1977). De patiënten dienden dan minder dan één standaarddeviatie van het "normale" gemiddelde af te wijken wilde van een gewenst (c.q. voldoende) resultaat gesproken kunnen worden. Deze methode zal o.a. ook in hoofdstuk 10 worden gebruikt. Een andere mogelijkheid is om met de patiënt *vooraf* de mate van gewenste verandering c.q. het eindresultaat vast te leggen, zoals dit bij de Goal Attainment Scaling (zie beneden) gebeurt. Uiteraard zal dikwijls een verschil van mening bestaan tussen door de patiënt gewenste en door de therapeut als haalbaar geziene verandering. Consensus hierover zal ook het therapeutisch proces zelf ten goede komen.

4.3 GEINDIVIDUALISEERDE DOELEVALUATIES.

Het motto van het gebruik van geïndividualiseerde doelevaluaties is goed verwoord door Davis (1973): "*It makes the difference between playing a football-game without or with goalposts and a scoring system*". Een toepasselijk citaat wordt reeds gevonden bij Aristoteles' *Metafysika* (ca. 330 v. Chr.). In het begin van boek 3 hanteert hij de vergelijking met de wandelaars die niet weten waarheen zij moeten gaan en

daarom niet zullen herkennen of ze het gezochte doel hebben bereikt of niet. (Vertaling: Seidel, 1982, p. 83).

Het principe van doelevauaties is even eenvoudig als voor de handliggend: *Vooraf bepalen wat de doelstelling is en achteraf nagaan of deze is bereikt*. Wat programma-evaluatie voor een heel programma of instelling is, is doelevauatie voor een bepaalde behandeling van een cliënt. Het is daarom ook niet verwonderlijk, dat deze evaluatie methode vooral binnen dit onderzoeksterrein ontwikkeld is. Het is eveneens een bewijs voor de scheiding der geesten tussen die van de onderzoekers en die van de evaluatoren (zie hoofdstuk 2) dat deze vorm van operationalisatie nauwelijks is doorgedrongen tot het psychotherapie-effect onderzoek. Over diverse gestandaardiseerde doel-evaluatie procedures die variëren in mate van specificering van de doelstellingen, is veelvuldig gepubliceerd. Onbekendheid kan hen derhalve niet worden toegeschreven. De belangrijkste methoden worden hier geresumeerd.

4.3.1 Target Complaints

De eenvoudigste procedure is de zgn. "*Target complaints*" (Battle e.a., 1966) of "*Target symptoms*" (Sloane, 1975). In Nederland is deze wel eens letterlijk vertaald met de minder fraaie term: doelklachten (Broekman en Schaap, 1984). Een geschikte vertaling is nog niet voorhanden, zodat hier de engelse term gehanteerd wordt. De werkwijze is als volgt: Bij de start van de behandeling formuleert de patiënt een, twee of drie klachten of problemen waarvoor hij hulp zoekt. Dit gebeurt door de patiënt zelf, zonder tussenkomst van de behandelaar, waardoor het theoretisch kader van de psychotherapeut de formulering niet kan beïnvloeden en zo bijv. vergelijking van patiënten tussen verschillende behandelaars onmogelijk zou maken (zie ook Lange, 1984). Bij beëindiging van de behandeling worden de klachten weer aan de patiënt voorgelegd met de vraag aan te geven of deze veranderd zijn en in welke mate: van geheel verbeterd tot (sterk) verslechterd.

De rationale van deze methode is: "*The target complaint concept is similar to the treatment criteria in other areas of medicine. The patient comes to a psysician with a series of complaints of symptoms; the alleviation of these complaints is the criterion of the efficacy of the treatment*" (Battle, 1966 p. 185, Kiesler, 1977 p.9). Onderzoek over deze methode toont aan dat de verbeteringsscores (1 = slechter, 2 = hetzelfde, 3 = een beetje beter, 4 = tamelijk veel beter, 5 = veel beter) correleren met andere effectmaten zoals globale verbetering beoordeeld door de patiënt en de therapeut. Ook de stabiliteit van de genoemde problemen blijkt hoog te zijn, inclusief de volgorde van belangrijkheid. Vergelijking met een standaard klachtenlijst toont aan dat "target complaints" wél differentiëren tussen een experimentele en een controle-groep en de klachtenlijst niet. De klachtenlijst kan nl. de belangrijkste klachten van de patiënt missen of te weinig

laten meetellen in de totaalscore, luidt de verklaring van de onderzoekers.

Er wordt een onderscheid gemaakt tussen de werkwijze waarbij de patiënt zelf zijn problemen formuleert of waarbij dit door een onafhankelijke onderzoeker wordt gedaan. Deze laatste methode biedt het voordeel van een grotere standaardisatie en een betere specificiteit van de klachten. Te vaag geformuleerde klachten zoals "ik wil me beter voelen" of "ik wil beter met anderen kunnen opschieten" kunnen onder invloed van de interviewer nader gepreciseerd worden. Uit onderzoek blijkt, dat bij het werken met interviewers de interbeoordelaarsbetrouwbaarheid groot is. Het nadeel is echter de grote arbeidsintensiviteit ervan. Bij het berekenen van een verbeteringsscore worden de beoordelingen over de verschillende klachten gemiddeld.

Aan het slot van zijn uitgebreide literatuuronderzoek naar de "target complaints" besluit Kiesler (1977) met de opmerking dat met deze evaluatiemethode bij voortdurende systematische toepassing informatie verkregen wordt *welke behandelingsstrategie voor welke problemen het meest effectief is*. Hiermee wordt een stap gezet om het rooster ("grid") in te vullen dat leidt tot het antwoord op de cruciale psychotherapie onderzoeksvraag: Welke behandelingen door welke therapeut zijn het meest effectief voor welke problemen van welke patiënten (p. 27); ofwel het multidimensionele onderzoeksdesign dat Paul (1967) reeds als utopisch beeld voor ogen had. Hieraan kan worden toegevoegd dat dit dan in de *klinische praktijk* zal moeten gebeuren, willen de resultaten bruikbaar zijn. De kloof tussen onderzoek en praktijk van de psychotherapie wordt dan gedicht vanuit de praktijkoever, ook in zijn operationalisatie van de effecten. De "Target Complaints"-methode kan hier een belangrijke bijdrage aan leveren.

4.3.2 Goal Attainment Scaling

Een tweede doelevaluatie methode die veel aandacht heeft gekregen, ook in Nederland, is de *Goal Attainment Scaling* (G.A.S.) (Kiresuk, 1973; Sherman, 1974; Kiresuk & Garwick, 1975; Seaburg & Gillespie, 1977a,b; Sherman 1977a,b; Kiresuk & Lund, 1978). Deze evaluatieprocedure is ontworpen door Kiresuk en Sherman (1968). De werkwijze is als volgt: In een interview met de patiënt worden de belangrijkste probleemgebieden geselecteerd. Deze zijn in eerste instantie globaal, zoals: relatie met partner, straatvrees, e.d. Per probleemgebied worden een of meer doelen gekozen die vervolgens in concrete termen moet worden vertaald. Dit worden de *indicatoren* genoemd. Een indicator voor straatvrees is bijv. de afgelegde afstand vanaf de eigen woning. Deze indicator moet specifiek zijn en bij voorkeur in de vorm van (voor anderen) waarneembaar gedrag. Van deze indicator worden vijf verschillende niveau's benoemd die corresponderen met de mate van realise-

ring van het doel: van veel minder dan verwacht via het verwachte niveau naar veel beter dan verwacht. Er moet ook vastgelegd worden wannéér c.q. gedurende welke periode en door wie de metingen verricht moeten worden. Al de verzamelde informatie wordt in een evaluatieformulier vermeld. Bij beëindiging van de behandeling en/of gedurende de follow-up wordt vervolgens per indicator nagegaan welk niveau bereikt is. Per formulier kunnen nu twee evaluatiescores worden berekend, de begin- en eindscore. De mate van verbetering kan worden bepaald door het verschil van deze twee. In een criterium georiënteerde situatie dient de eindevaluatie voor iedere indicator (minimaal) het verwachte niveau te bereiken hetgeen correspondeert met een totaalscore van 50 voor het hele formulier. Afwijkingen hiervan kunnen dan in relatie tot dit criterium geëvalueerd worden. (Zie ook: Melief e.a., 1979).

In Amerika heeft deze methode zeer veel aandacht gekregen en is met overheidssteun verdere ontwikkeling en verspreiding ervan mogelijk geweest. In 1976 was reeds een literatuurlijst van meer dan 200 onderzoeksverslagen beschikbaar (Garwick e.a., 1976). In diverse onderzoeksprojecten is de betrouwbaarheid ervan nagegaan (Kiesler, 1977, p. 91 e.v.). Deze is doorgaans zeer hoog: $r = .77$ tot $.96$. Ook wanneer verschillende onderzoekers dezelfde patiënt interviewen en hiervan een G.A.S. construeren, is de overeenstemming groot. In diverse validiteitsstudies zijn de G.A.S.-scores vergeleken met andere effect-metingen zoals de M.M.P.I., Brief Psychiatric Rating Scale, e.d. Het is niet verwonderlijk dat de correlatie met deze metingen niet hoog is: $.20$ tot $.52$. G.A.S. is juist ontwikkeld, omdat de gestandaardiseerde en uniforme operationalisaties ontoereikend zijn om therapie-effect te meten.

Met name in programma-evaluatie projecten is de G.A.S. veel gebruikt en ontbreekt dan ook niet in de grote handboeken over deze materie (Attkisson & Hargreaves, 1978; Coursey, 1977). Bij het psychotherapie-effect onderzoek wordt deze evaluatie-methodiek hoegenaamd niet aangetroffen. Soudijn (1982) noemt in zijn boek "De kwaliteit van de psychotherapie" G.A.S. als *diagnostisch instrument* en meent dat "... het in dat opzicht vreemd aan doet". (p. 52). Maar G.A.S. is ook niet als diagnosticum bedoeld. De kritiek die hij verder opsomt, is wel relevant en vaker geuit: de vrijheid van keuze (welke problemen, welke indicatoren, welke verwachtingen) is inderdaad zeer groot en foute inschattingen kunnen licht gemaakt worden. Uit onderzoek blijkt evenwel dat dit meevalt, wanneer gebruik wordt gemaakt van goed gemotiveerde en ervaren G.A.S.-constructeurs. En de doorsnee therapeut is dit niet. Dit laatste wordt ook door Kiesler (1977) als belangrijkste nadeel genoemd: *"Perhaps the most frequently pointed to difficulty concerns the expense and efforts to initiate and continue its application in mental health agencies"* (p. 118). Diverse auteurs worden geciteerd die van mening zijn, dat alleen met zeer gemotiveerde therapeuten die bereid zijn de nodige tijd hierin te investeren, het mogelijk zal zijn

deze procedure in te voeren, waarna zij blijvend dienen te worden ondersteund.

Dit nadeel werd ook ervaren gedurende een proefproject in 1978 (Sijben, 1981, p. 10). De bewerkelijkheid van het G.A.S.-formulier, waarbij diverse stadia van verwacht eindresultaat geëxpliciteerd moeten worden, was mede debet aan het mislukken van deze proef (zie ook hfst. 6).

Het G.A.S.-formulier heeft ook nog een *psychologisch* (en *praktisch*) bezwaar: Door de lay-out krijgt de invuller het idee heel veel informatie over een patiënt in een aantal kleine hokjes te moeten persen. Om deze redenen werd in het onderzoek dat in deel B wordt beschreven, de G.A.S.-procedure ingeperkt. Ter verduidelijking werd de naam veranderd in *Individuele Doel Evaluatie (I.D.E.)* en *Individuele Doel Evaluatie Meting (I.D.E.M.)* bij het bepalen van het resultaat. Deze I.D.E.(M.)-procedure verloopt als volgt: de therapeut (of externe beoordelaar) benoemt een probleemgebied of een globaal doel, specificceert hiervan een of meer indicatoren en geeft per indicator aan wat het beginpunt is en wat het streefniveau. Afwijkingen van dit streefgetal geeft aan de mate van het welslagen van de behandeling. Middels tussentijdse evaluaties kan het veranderingsproces gevolgd worden. De procedure is minder arbeidsintensief en is gebruikersvriendelijker. De *creativiteit* om met name goede indicatoren te construeren en de *discipline* om voor iedere patiënt weer een nieuwe I.D.E. te maken en periodiek te evalueren, blijven potentiële moeilijkheden (zie hfst. 10.2).

Wellicht is dit ook de reden dat G.A.S. in Nederland slechts op beperkte schaal is gebruikt. In de jaren zeventig is in het Maandblad Geestelijke Volksgezondheid een artikelenreeks verschenen over deze evaluatie methodiek, voornamelijk m.b.t. de methodologische en statistische thema's (Van der Hart & Lange, 1972; Bensing & Dekker, 1973; Schmidt, 1974; Van der Veer, 1974; De Wit, 1975). De onderzoeken waarin G.A.S. is gebruikt, zijn minder in aantal (Beenackers, 1976; Corveleyn, 1978; Veugelers, 1978; Sijben, 1979d). Van de in 1979 verschenen handleiding "Doelen stellen en evalueren" (Melief e.a., 1979) zijn slechts 584 exemplaren verspreid of verkocht en is in 1986 uit de handel genomen. Door auteurs en uitgeverij was de verkoopmogelijkheid indertijd hoger ingeschat.

4.3.3 Andere doel-evaluatie methoden.

Van andere doelevaluatie methoden verdient *Automated Tri-informant Goal Oriented Note (ATGON)* het om gememoreerd te worden (Wilson & Ellis, z.j.; Wilson, 1973a,b; 1974, 1975, 1977) Deze in Fort Logan Hospital, Denver, USA opgezette evaluatiemethode bestaat uit een geautomatiseerde procedure, waarin door drie informanten (patiënt, hulpverlener en significante ander) gegevens worden verstrekt. Het systeem bevat globale doelen, die door de drie informanten worden beoor-

deeld in termen van meer of minder veranderd. Ook de gebruikte therapeutische technieken worden geëvalueerd. Met veel overheidssteun is dit systeem in 1967 gestart. Bij een bezoek in 1979 bleek het project ter ziele: de subsidie was gestopt, nog één medewerker verzorgde het archief en verzond materiaal waarom gevraagd werd. Toch kan dit project qua methodiek een voorbeeld zijn om doelevaluaties te koppelen aan therapeutische strategieën en zo bij langdurig gebruik in een praktijksetting waardevolle informatie verschaffen over de interactie tussen problemen (doelen) en middelen.

Andere doel-evaluatie methoden worden beschreven in Melief e.a. (1979, p.149 e.v.) en Sijben (1979a zie bijv. Bonstedt, 1973). Deze hebben weinig algemene bekendheid gekregen. Het geautomatiseerde systeem van Meldman (1976) is evenmin op grotere schaal verspreid. Het was toentertijd zijn tijd (te) ver vooruit.

Dit besluit deze beschrijving van geïndividualiseerde doelevaluaties. De twee methoden: target complaints en G.A.S. of I.D.E.(M.) kunnen worden beschouwd als belangrijke instrumenten om in de praktijk de behandelingseffecten te meten.

4.4 PSYCHOLOGISCHE TESTS.

In par. 4.2 is reeds beschreven dat de meest gebruikte evaluatie-instrumenten psychologische tests zijn. Door sommigen worden deze zelfs de belangrijkste genoemd. In grote meerderheid betreffen dit *vragenlijsten* die de patiënt zelf vóór en ná de behandeling invult. De verschillen tussen deze twee meetmomenten moeten het resultaat van de behandeling aangeven. De voordelen van deze instrumenten zijn de aange-toonde of vermeende psychometrische kwaliteiten zoals betrouwbaarheid en validiteit, het beschikbaar zijn van vergelijkende normgroepen en vooral de gemakkelijke afname en scoring.

De keuze voor een bepaalde test dient primair bepaald te worden door het te meten *construct*, of althans dat wat de test of zijn maker pretendeert te meten. Voor de evaluator van therapie-effecten bestaat er een grote diversiteit aan meetinstrumenten. Deze hebben betrekking op algemene aspecten als persoonlijkheidsvariabelen of op specifieke problemen zoals angst, depressie, dwang, lichamelijke klachten, stemmingen (mood scales), e.d. Met name bij de evaluatie van gedragstherapie wordt vaak van deze specifieke lijsten gebruik gemaakt (Hersen en Bellack, 1976, Sijben e.a. 1979e). Sommige zijn zelfs binnen deze therapeutische stroming ontwikkeld, zoals de Fear Survey Schedule (Wolpe & Lang, 1964), en diverse assertiviteitsschalen (Brinkman, 1978). In Nederland worden met name gebruikt: Amsterdamse Biografische Vragenlijst (Wilde, 1963) en Nederlandse Persoonlijkheidsvragenlijst (Luteijn, 1969) voor het meten van persoonlijkheidsvariabelen. De Minnesota Multiphasic Personality Inventory

(Hathaway, 1963; Nuttin & Beuten, 1969), welke door Bergin & Lambert (1978) wordt aanbevolen, is weinig bruikbaar. De test is te lang (550 vragen) en is in het Nederlandse taalgebied onvoldoende gevalideerd en genormeerd. Voor het meten van specifieke problemen worden met name gebruikt:

- voor algemene angst: Taylor Manifest Anxiety Scale (Taylor, 1953; Bendig, 1956)
- voor sociale angst: Willems Sociale Angstschaal (Willems e.a. 1973)
- voor fobische angst: Fear Survey Schedule ((Wolpe & Lang, 1964)
- voor assertiviteit: Brinkman assertiviteitsschaal, Schaal voor Interpersoonlijk Gedrag (Arrindell e.a., 1986)
- voor dwanggedachten /-handelingen: Inventarisatie dagelijkse bezigheden (Kraaimaat & van Dam-Baggen, 1976)
- voor depressie: Zung depressieschaal (Zung, 1965), VROPSOM en de Beck depression inventory (Hoevenaars & Van Son, 1984).
- voor lichamelijke klachten: Brinkman klachtenlijst (Brinkman 1978), Symptom Check List-90 (Derogatis, 1973)

Volgens de Commissie Test Aangelegenheden Nederland (COTAN) voldoen alleen de persoonlijkheidsvragenlijsten A.B.V. en N.P.V. aan alle door hen gehanteerde beoordelingsnormen. Alle overige tests hebben een of meer onvoldoendes of zijn zelfs niet in hun uitgave opgenomen. In dat opzicht kleven ook aan de vaak gebruikte vragenlijsten de nodige bezwaren, met name op het terrein waarom ze juist in therapie-effect onderzoek worden aanbevolen: hun psychometrische kwaliteiten. Recent onderzoek met de F.S.S. (Arrindell, 1980; Arrindell & Van de Ende, 1986; voor dit laatste artikel zijn de gegevens uit het hierna volgend onderzoek gebruikt) en de SCL-90 (Arrindell & Ettema, 1978, 1981) kan de beoordeling van deze tests wellicht veranderen.

4.5 SUBJECTIEVE BEOORDELINGEN.

Wellicht de meeste gebruikte meetinstrumenten zijn vragenlijsten die *globale* verbetering trachten te meten. Hiertoe kunnen ook de zgn. cliënt-satisfactielijsten gerekend worden. Een voorbeeld van een globale verbeteringslijst is de Global Assessment Scale (Endecott e.a., 1976; Hargreaves & Attkisson, 1978, p. 306 e.v.), waarmee globaal functioneren van de patiënt op een 100-puntsschaal kan worden beoordeeld door de behandelaar. Bij zelfbeoordelingsprocedures wordt deze schaal ingekort tot 5 of 6 antwoordcategorieën. Vragen als: Zijn uw klachten verbeterd, hoe voelt u zich, hoe beoordeelt u uw dagelijks functioneren, enz. zijn alle globale beoordelingen van het effect van de behandeling. De *schijnvaliditeit* van deze lijsten is groot. Ze genieten (daarom?) een grote voorkeur bij evaluatoren. Het nadeel is de grote reactiviteit van deze metingen; ze zijn gevoelig voor "hello-goodbye

effecten". Hieronder wordt verstaan dat de invuller de neiging heeft tot het geven van een te negatieve voorstelling van zaken aan het begin van een behandeling, en een het einde ervan een positieve vertekening van het effect zal geven (Soudijn, 1976, p. 68). Evaluatie van effecten uitsluitend op basis van dergelijke vragen wordt doorgaans als te mager bestempeld.

Met name in Amerika is een vrij omvangrijk gebruik van cliënt-satisfactielijsten gemaakt. Dit zijn eveneens globale beoordelingen maar hebben meer betrekking op het programma in zijn geheel dan op de resultaten ervan. De ideeën over het meten van de tevredenheid over een behandeling zijn afkomstig uit het commerciële consumentenonderzoek. Hierin zijn drie modellen ontwikkeld (Pascoe, 1983a,b):

- het *contrastmodel* dat zich richt op het ervaren verschil tussen verwachte en ervaren dienstverlening;
- het *assimilatiemodel* waarin wordt aangenomen dat een geringe discrepantie tussen verwachting en geleverde prestatie spanning oproept bij de klant en dat deze leidt tot een verandering in de waarneming van de prestatie;
- het *assimilatie-contrastmodel*. Dit postuleert een gedifferentieerde reactie op de aard en de mate van verschil tussen verwachting en prestatie. In een gebied rond de verwachte waarde zullen verwachtingen die afwijken van de prestatie deze herwaarderden, zodat het verschil kleiner wordt (assimilatie). Daarbuiten zal het contrastmodel werkzaam zijn.

Over de *houdbaarheid* van deze modellen lopen de meningen uiteen. Dat processen van assimilatie en contrast een rol zullen spelen bij de beoordeling van geleverde diensten, lijkt onbetwistbaar.

Het onderzoek naar satisfactie in de geestelijke gezondheidszorg is van recentere datum. Doorgaans worden weinig gestandaardiseerde, ad hoc opgestelde lijstjes gebruikt. Een speciaal nummer van *Evaluation and Program Planning* (1983) geeft in 18 artikelen, samen ruim 200 pagina's, een overzicht van de stand van zaken op dit terrein. Gepleit wordt voor een *multidimensionele benadering* door diverse aspecten van de tevredenheid te traceren (Pascoe, 1983). Lebow (1983a,b) geeft een overzicht van 63 onderzoeken naar cliënt-satisfactie. Hieruit blijkt, dat gemiddeld slechts 57 % de lijst invult, variërend van 19% tot 100%. Over het algemeen zijn de cliënten tevreden over de geleverde service; klinische patiënten geven een gemiddelde van 76% tevredenheid aan, hetgeen het laagste is. Bepleit wordt om aan het percentage ontevreden patiënten extra aandacht te wijden. De correlatie tussen tevredenheid en de globale beoordeling van het resultaat blijkt hoog te zijn. In een meta-analytische studie van Lehman en Zastowny (1983) wordt getracht normen op te stellen voor cliënt-satisfactie op basis van 59 onderzoeken. Per type voorziening ("inpatient/outpatient", "chronic/non-chronic" en "conventional/innovative") worden voorlopige normen gegeven. Voor "inpatient, non-chronic conventional set-

tings" is een gemiddelde Z-waarde van 1.62 met een standaarddeviatie van 1.14 berekend, voor een vijfpunts-schaal. De betekenis van een dergelijke norm is *triviaal*. Er zijn verschillende lijsten gebruikt en de resultaten per kliniek zijn zeer uiteenlopend. Een gemiddelde score voor een evaluatie-aspect dat in principe criterium georiënteerd moet worden beoordeeld, is van weinig nut. De aanbeveling voor een "cumulative, national data base on patient satisfaction ... to further refine these norms" (p. 265) dient zelf een grotere verfijning dan in het artikel gegeven wordt. Een pleidooi voor meer *standaardisatie* zou eveneens gewenst zijn.

Een lijst die hier de meeste aanspraak op maakt, is de *Client Satisfaction Questionnaire (C.S.Q)*. Deze lijst is op het Langley Porter Institute San Francisco U.S.A. ontwikkeld (Hargreaves & Attkisson, 1978; Nguyen e.a., 1983). Aanvankelijk werden negen dimensies aangenomen op basis van literatuuronderzoek. Dit waren: de fysieke faciliteiten zoals het gebouw, de bereikbaarheid e.d., de steun van de medewerkers, de soort behandeling, de bekwaamheid van de behandelaar, de kwaliteit, de duur en kwantiteit van de service, het resultaat van de behandeling en de algehele tevredenheid. Via opeenvolgende analysetechnieken ontstond een 31-item vragenlijst. Uit factoranalyse bleek, dat deze lijst een algemene factor bleek te bevatten, bestaande uit 8 vragen. Dit mondde vervolgens uit in de Client Satisfaction Questionnaire met een coëfficiënt alpha van .93. In diverse onderzoeksprojecten is deze lijst gebruikt, waarbij blijkt dat er een significante correlatie bestaat met de resultaten van de behandeling en diverse standaardvragenlijsten zoals de SCL-90. In contrast tot de variëteit aan zelf ontworpen lijstjes is de C.S.Q. een goed ontwikkelde en onderzochte satisfactieschaal die verdient op grote schaal te worden gebruikt. Er bestaan momenteel zowel een Nederlandse (Sijben, 1981; De Brey, 1983) als een Spaanse versie (Roberts & Attkisson, 1983).

Een laatste aspect met betrekking tot de meting van de satisfactie is de *anonimiteit* van de gegevens. In het onderzoek van Soelling en Newell (1983) werd de C.S.Q. in vier afnamecondities gebruikt: al dan niet vertrouwelijk en wel of niet met een extra verzoek om toch vooral eerlijk te antwoorden. Deze laatste conditie gaf geen verschil in resultaten. Het anoniem invullen van de lijst bleek op alle vragen tot een significant lagere tevredenheidsscore te leiden. Vrees voor repercussie wanneer nog eens van dit instituut gebruik gemaakt zou worden, zal de verklaring zijn voor dit verschijnsel ("*You do not complain about your surgeon's skills while on the operating table*", p. 332). De aanbeveling uit dit onderzoek dat de vragenlijst anoniem dient te worden ingevuld, moet worden opgevolgd.

Unaniem worden steeds hoge tevredenheidsscores gerapporteerd. Aangezien het vermoeden bestaat dat dit een te positieve vertekening is, zijn methoden uitgeprobeerd om dit te veranderen (Pascoe en Attkisson, 1983; Attkisson, Roberts en Pascoe, 1983). Deze zijn echter slechts ten dele succesvol. Eén methode is niet gebruikt: *het vergroten van het aantal*

schaalwaarden. De C.S.Q. hanteert vier meetpunten zonder een neutraal midden. Een erg scheve verdeling kan verbreed worden door de antwoordmogelijkheden te vergroten. In het onderzoek in deel B zal worden nagegaan of dit een verandering van de verdeling te zien geeft.

4.6 CONCLUSIE.

Het psychotherapie-effect onderzoek heeft weinig resultaten voor de praktijk opgeleverd, luidt reeds decennia lang de klacht. Een oorzaak is ook gelegen in de gebruikte effectmetingen. Deze zijn doorgaans weinig specifiek, niet gericht op de *doelstellingen* van de patiënt. Vanuit de praktijk van de programma-evaluatie zijn diverse doel-evaluatie methoden ontwikkeld welke ook goed bruikbaar zijn voor de evaluatie van behandelingsresultaten. Dergelijke doelmetingen, samen met gestandaardiseerde vragenlijsten lijken geschikt om in de praktijk de effecten van de psychotherapie te evalueren.

Een dergelijke benadering voldoet ook aan het door Campbell & Fiske (1959) gepropageerde "*multitrait-multimethod design*", en dat met name voor psychotherapie onderzoek zeer waardevol is. In een langlopend project in een praktijksetting zou met deze verschillende instrumenten ervaring moeten worden opgedaan. Zo kan naast programma-evaluatieve gegevens (*registratie*) ook een bijdrage worden geleverd aan de *verklaring* voor de al dan niet opgetreden veranderingen. De *tevredenheid* van de patiënt is een maat die in ieder geval thuis hoort bij de evaluatie van een behandeling. De slogan: *bent u tevreden zeg het anderen, hebt u klachten zeg het ons*, dient standaard ook voor de geestelijke gezondheidszorg te gelden. Middels de Client Satisfaction Questionnaire wordt de patiënt hiertoe in de gelegenheid gesteld.

Programma-evaluatie in praktijk

*Het was, voor ik met het schrijven ervan begon,
mijn bedoeling geweest er alles in op te nemen
wat ik wist over het wezen van de stoffelijke dingen,
maar, zoals ook kunstschilders in een vlak schildery
niet alle kanten van een voorwerp
even goed kunnen afbeelden,
en er slechts een van de belangrijkste uitkiezen
waarop zij het licht laten vallen
– de andere kanten laten zij in de schaduw
zodat zij hooguit zydelings aan bod komen –,
zo vreesde ik niet alles te kunnen zeggen
wat ik wilde,*

DESCARTES, Over de methode (1637)

Hoofdstuk 5:

Kliniek Overwaal

"De doelen en kenmerken van een programma of van de activiteiten die dienen te worden geëvalueerd, moeten zo nauwkeurig mogelijk worden gespecificeerd."
le richtlijn voor programma-evaluatie.
(De Brey, 1981)

5.1 INLEIDING.

Programma-evaluatie betekent het *evalueren van doelstellingen*. Deze doelstellingen volgen o.a. uit de beoogde behandeling en werkwijze, de patiëntenpopulatie en de therapeutische strategie. Daarom dient het werkterrein uitgebreid beschreven en de behandelingsprocedure geëxpliciteerd te worden. Dit leidt ook tot een beter begrip voor de opzet van het programma-evaluatieproject (zie hoofdstuk 6); de evaluatiemomenten en -procedures sluiten aan bij de werkwijze. Een goed inzicht in deze relatie kan tevens behulpzaam zijn bij het vertalen van de ervaringen van dit project voor andere instellingen.

5.2 DOELSTELLING.

Kliniek Overwaal is een categoriale instelling, omschreven als een *kliniek voor psychotherapie*. In het jaarverslag van Overwaal staat de volgende beknopte doelstelling vermeld:

In Overwaal worden patiënten opgenomen met psycho-sociale problemen, die niet of niet voldoende behandeld kunnen worden bij de ambulante voorzieningen voor hulpverlening. De behandeling is afgestemd op de problematiek van elke patiënt afzonderlijk en bestaat uit diverse vormen van psychotherapie. Het behandelingsdoel is het bereiken van zodanige veranderingen dat de patiënt zonder verdere hulp zijn leven kan inrichten ofwel met vrucht (weer) gebruik kan maken van ambulante voorzieningen.

(Overwaal jaarverslag 1977)

In 1983 werd deze gewijzigd en werd o.a. de passage: "zonder verdere hulp zijn leven kan inrichten", verwijderd.

Deze doelstelling maakt duidelijk dat Overwaal gezien moet worden als een tussenfase in een langer durend behandelingsproces. Na een mogelijk ambulante behandeling die om uiteenlopende redenen niet tot succes leidt, volgt een intensieve klinische behandeling, waarna een ambulante hulpverlener het therapieproces voortzet. Dit is het ideaaltypische behandelingsverloop. Hiermee samenhangend is het doel dat de opname-duur *kort* dient te zijn. Gedacht wordt dan aan vier à vijf maanden. Naast een therapeutisch doel om de behandelingsduur zo kort mogelijk te houden, is dit ook een nadrukkelijke wens van de financiers (COTG). Het budgetaire systeem is erop gebaseerd dat voor kortdurende behandelingen tot max. 6 maanden een hogere verpleegprijs kan worden berekenen - de zgn. A1 patiënten - dan wanneer de behandelingsduur zes maanden of meer bedraagt, aangeduid met A2 patiënten. De redenering is dat een korte behandeling arbeidsintensiever en dus duurder kan en mag zijn dan een langerdurende behandeling. Overwaal acht zichzelf een "A1-kliniek" en is ook jarenlang door het COTG als zodanig financieel beoordeeld. Dit is een externe oorzaak om de behandelingsduur binnen de A1-norm te houden. Hoe bovenbeschreven doelstelling in praktijk vorm wordt gegeven, wordt in Par. 5.4 en 5.5 verder uitgewerkt met een beschrijving van de werkwijze en de behandelingsfasen.

5.3 VOORGESCHIEDENIS.

De geschiedenis van Overwaal is nog vrij jong en gaat terug tot 1971. In dat jaar werden onder leiding van Prof.Dr. S.J. Nijdam, hoogleraar psychiatrie van het Sint Radboudziekenhuis, de uitgangspunten voor een nieuw op te richten kliniek uitgewerkt. Deze punten waren:

- Overwaal wordt niet bestemd voor het hele scala van psychiatrische patiënten, maar voor een speciale categorie, namelijk *curabele gevallen uit de lichtere psychiatrie* met uitsluiting van kinderen, pubers en bejaarden.
- Overwaal participeert in de *wetenschapsbeoefening* en in het onderwijs van de universiteit (de kosten hieraan verbonden vallen buiten de verpleegprijs).
- De behandeling zal een *multidisciplinair* karakter hebben.
- De opzet zal zeker in de aanvang niet te ver afwijken van de *klassieke kliniekstructuur*. Zo werd bijvoorbeeld uitdrukkelijk gesteld dat niet begonnen wordt aan een therapeutische gemeenschap; wel zou, meer dan in de klassieke kliniek, aandacht worden gegeven aan de dagelijkse begeleiding van de patiënt (therapeutisch milieu). Deze begeleiding echter staat ten dienste van de behandeling die vanuit verschillende disciplines naargelang van de behoefte op elke patiënt wordt toegepast.
- Het psychiatrisch ziekenhuis heeft een dimensie meer dan de andere klinische specialismen, namelijk de voortdurende zorg voor de *gehele woon- en werksituatie* van de aan zijn zorg toevertrouwde patiënten. (Overwaal jaarverslag 1974)

De organisatiestructuur was complex. "Huize Overwaal", zoals dit toentertijd werd genoemd, ressorteerde onder de Sint Martinusstichting, Samivoz te 's-Hertogenbosch, welke zich tot taak had gesteld een "*neurosenkliniek*" op te richten. Functioneel sloot de kliniek nauw aan bij de afd. psychiatrie van het Sint Radboudziekenhuis. Er werd een samenwerkingsverband aangegaan: de hoogleraar psychiatrie werd hoofd van kliniek Overwaal, er werden assistenten-psychiatrie opgeleid en de bedden van Overwaal telden mee bij de afd. psychiatrie die zelf te weinig bedden had om als opleidingsplaats erkend te worden.

In 1972 werd de *kliniek-erkenning* verkregen en in 1973 werd de verpleegprijs vastgesteld, waarna met name het psychotherapeutisch team kon worden gecompleteerd. Naast de reeds aanwezige chef de clinique (psycholoog), twee assistentpsychiaters en B-verpleegkundigen, werden een psychiater, arts-psychotherapeut en psychologen aangetrokken. Intern wordt het jaar 1974 gezien als startjaar van kliniek Overwaal omdat toen de psychotherapeutische behandeling conform de doelstellingen kon beginnen. De patiëntenpopulatie zal dan ook vanaf dat jaar beschreven worden (zie hfst. 7.2).

5.4 WERKWIJZE.

Overwaal staat voornamelijk bekend als een kliniek voor *gedragstherapie*. In dit therapiemodel zijn of worden alle psychotherapeuten opgeleid en de organisatie van de behandeling is hier sterk op georiënteerd. Volgens gedragstherapeutische principes dienen behandelingsdoelen en werkwijze zo goed mogelijk te worden *geëxpliciteerd*. De behandeling wordt uitgevoerd door drie groepen van therapeuten:

1. *Psychotherapeuten*. Diverse disciplines zijn werkzaam (geweest) als psycho-therapeut: psychiater, psychologen, andragoge, arts-psychotherapeute en assistent-psychiaters.
2. *Sociotherapeuten*. Aanvankelijk waren de sociotherapeuten ofwel B-verpleegkundigen ofwel afgestudeerden van de Sociale Academie. De laatste jaren prevaleren B-verpleegkundigen voor deze functie.
3. *Overige therapeuten*. Hieronder worden begrepen: creatieve, arbeids-, bewegings-, drama- en expressietherapeuten.

De werkwijze is sterk *individueel* gericht. De psychotherapeut draagt de hoofdverantwoordelijkheid van de behandeling. Hij/zij stelt het behandelingsplan op, voert diverse onderdelen van dit plan zelf uit zoals met name de individuele psychotherapie en heeft de regie van de behandelingsactiviteiten van de overige therapeuten.

Een psychotherapeut vormt samen met een vaste sociotherapeut en ca. vijf patiënten een zgn. *werkgroep*. Deze twee therapeuten hebben intensief overleg over de behandeling van

"hun" patiënten; de sociotherapeut functioneert in deze als een co-therapeut. Een patiënt wordt bij opname aan een bepaalde psycho- en sociotherapeut toegewezen, hetzij willekeurig hetzij geïndiceerd, en blijft doorgaans tot het ontslag onder hun behandeling. De overige therapeuten zijn werkzaam op de afdeling Creatieve therapie en Bewegingstherapie. Op deze afdelingen zijn verschillende programma's: groeps- en individuele therapieën waar patiënten op indicatie van de psychotherapeut en conform het behandelingsplan aan deelnemen.

Naast de individuele therapie van de eigen psycho- en sociotherapeut en de therapieën van de afd. Creatieve en Bewegingstherapie worden door psycho- en sociotherapeuten ook groepstherapieën verzorgd voor alle patiënten voor wie dit is geïndiceerd. Dit zijn o.a. assertieve training, ontspanningsoefeningen, communicatie-oefeningen (zgn. praatgroep).

5.5 DE BEHANDELINGSFASEN.

Het therapeutisch functioneren is in enkele fasen te onderscheiden: intake, opname en behandelingsplan, behandeling, ontslag en nazorg.

5.5.1 Het intake-proces.

De *aanmelding* geschiedt schriftelijk en soms telefonisch bij de chef de clinique. Deze aanmelding wordt besproken in een zgn. intake-team, bestaande uit chef de clinique, enkele psychotherapeuten en een secretaresse voor het noteren van afspraken e.d. Onder geleide van de gegevens van de verwijzer kan het intake-team besluiten een intake gesprek met de aangemelde patiënt te houden. Dit gesprek vindt plaats met een van de psychotherapeuten van het intake-team. Hiervan wordt een verslag gemaakt en bij een volgende intake-vergadering ingebracht. Het intake-team neemt vervolgens de beslissing tot opname of afwijzing. Bij onduidelijkheid is een tweede gesprek mogelijk met een andere therapeut. Als de patiënt wordt aangenomen, wordt dit aan de verwijzer en aan de patiënt meegedeeld. Deze krijgt dan tevens een informatie-stencil en een indicatie over wanneer hij zal worden opgenomen, hetgeen afhankelijk is van de lengte van de wachtlijst.

De beslissing wel of niet opnemen is zowel individueel voor de betrokken patiënt als institutioneel van groot belang. Wanneer door de intaker c.q. het intake-team een verkeerde inschatting wordt gemaakt van de problematiek van een aangemelde patiënt, bijv. dat betreffende problematiek niet binnen de behandelingsdoelstellingen van Overwaal valt qua ernst (bijv. psychoticisme of narcomanie) of qua geschatte behandelingsduur (veel langere therapie noodzakelijk dan ca. vijf maanden), dan zal dit het instituutsresultaat in nega-

tieve zin beïnvloeden. Daarentegen zullen de doelstellingen gemakkelijker gerealiseerd kunnen worden, wanneer patiënten opgenomen worden wier problematiek niet ernstig genoeg is om een klinische behandeling te rechtvaardigen. De kans dat deze laatste onterechte opname plaatsvindt is geringer dan de eerste mogelijkheid, omdat er voor a.s. patiënten vaak een drempel is voor een klinische behandeling (stichmatisering) en verwijzers evenmin niet te snel zullen aanmelden maar eerder en soms te lang proberen de problematiek zelf ambulant te behandelen.

Het intakeproces kan danook als *eerste evaluatie-moment* worden beschouwd. Op Overwaal worden de intake-indicaties op een 16 items tellende lijst beoordeeld. De intaker geeft op een vijfpuntsschaal aan in hoeverre de afzonderlijke indicaties van toepassing zijn. Vijf betekent dat geheel aan de indicatie voldaan wordt, een drie is twijfelachtig en een één betekent dat op dit onderdeel de patiënt in het geheel niet aan de doelstelling van Overwaal voldoet.

5.5.2 Van opname tot behandelingsplan.

Bij opname wordt een patiënt aan een bepaalde psychotherapeut toegewezen. Hij komt in een vaste groep van patiënten met een psycho- en sociotherapeut; deze groep wordt de "*werk-groep*" genoemd. Zoals gesteld, is de behandeling sterk individueel gericht. De eerste drie weken worden vooral gekenmerkt als kennismakings-, observatie- en informatie-periode. De eigen psychotherapeut stelt in die periode een *behandelingsplan* op, gebaseerd op informatie verkregen uit de individuele gesprekken van hemzelf en de sociotherapeut, de verwijsbrief en observatiegegevens van de overige therapeuten, o.a. verkregen uit diverse kennismakingsgroepen. Dit behandelingsplan moet binnen vier weken in de teamvergadering, een vergadering van alle therapeuten, worden ingebracht. Het wordt in grote lijnen opgesteld volgens een vaste structuur, welke reeds in 1975 werd gehanteerd (Overwaal, 1975) en later werd aangepast volgens het model van Brinkman (Brinkman, 1978, Overwaal, 1977). De te beschrijven informatie is:

1. Kennismaking, informatie, anamnese.
2. Probleembeschrijvingen.
3. Concretisering.
4. Baseline metingen.
5. Topografische en functionele analyse.
6. Doelstellingen.
7. Therapieplan.
8. Technieken en procedures.

Niet alle onderdelen worden even uitgebreid vermeld. Met name 4 en in minder sterke mate 3 en 6 worden hoogstens slechts summier aangeduid. Persoonlijke voorkeur van de opsteller en aard van de problematiek (o.a. operationaliseerbaarheid) speelt hier natuurlijk een rol.

Dit plan wordt voorgelezen aan de collega's die het kunnen becommentariëren, waarna het verder in praktijk wordt gebracht. Ook de patiënt krijgt hiervan een afschrift. Door- gaans is de therapeut reeds ten dele gestart met een behande- ling conform het ingebrachte plan, eventueel met wijzigingen zet hij dit voort.

5.5.3 De behandeling.

Het belangrijkste therapeutische middel, het *individuele gesprek* tussen psychotherapeut en patiënt vindt in het alge- meen dagelijks plaats. Daarnaast is een uitgebreid scala van therapeutische technieken voorhanden, die door de eigen psycho- of sociotherapeut of andere therapeuten worden uitge- voerd. In tabel 5.1 worden de diverse technieken vermeld, welke anno 1984 zijn gebruikt.

Tabel 5.1 Gebruikte therapeutische technieken/methoden per type therapeut

eigen psychotherapeut	psycho-/ sociotherapeut	andere therapeuten
ind. gesprekken *)	assertieve training	Kreatieve therapie
groepsgesprekken	praatgroep	- individueel
relatie/gezins- gesprekken	alg.dagelijks leven	- groep
imaginaire con- frontatie	ontspanningsgroep	- kursussen
psychodrama		Bewegingstherapie
		- individueel
		- groep
		- massage groep
		- sport
		- conditietraining

*) In de individuele gesprekken worden uiteraard tal van specifieke (gedrags)-therapeutische technieken gebruikt zoals flooding, rollenspelen, cognitieve herstructure- ring, counseling, etc.

Naast deze *psychotherapeutische benadering* is iedere patiënt ook aan een arts toegewezen. Deze verricht bij opname een lichamelijk onderzoek en wordt gedurende de behandeling geconsulteerd voor lichamelijke klachten en, zo nodig in overleg met de psychotherapeut, voor het uitschrijven van psychofarmaca. De psychotherapeut heeft de regie van de uit- voering van het behandelingsplan. Hij heeft regelmatig over- leg met de overige therapeuten, welke eveneens onderdelen uitvoeren zoals de kreatieve en bewegingstherapeuten, die eveneens bij de behandeling participeren.

De voortgang van de behandeling wordt eens per maand in het team geëvalueerd. De psychotherapeut houdt doorgaans hieraan voorafgaand een *evaluatiegesprek* met de patiënt over de resultaten die tot dan geboekt zijn en andere elementen van de behandeling, zoals motivatie, e.d. Een bij voorkeur schriftelijke rapportage van dit gesprek, de eigen indrukken van de therapeut alsmede een evaluatie van de behandelingsdoelen worden in de teamvergadering ingebracht en besproken. Sinds 1983 duurt een behandeling in eerste instantie vier maanden. Indien echter blijkt dat deze periode te kort is, kan de psychotherapeut verlenging aanvragen voor een of twee maanden. Het aantal verlengingen is in feite niet beperkt, maar bij een te lange opnameduur zal de druk vanuit het team en het management (chef de clinique) steeds groter worden om de opname te beëindigen, eventueel met doorverwijzing naar een andere klinische setting waar een langere opnameduur wel mogelijk is.

5.5.4 Het ontslag.

De meest voorkomende reden voor ontslag zal zijn, dat de behandeling zo ver gevorderd is dat ambulante hulp (weer) toereikend zal kunnen zijn. In een klein aantal situaties zal doorverwijzing naar een andere kliniek geïndiceerd zijn, terwijl ook een eenzijdige beëindiging van de opname door de patiënt of therapeut zal voorkomen om uiteenlopende redenen: onenigheid over de behandeling, weer naar huis willen van de patiënt bijv. voor de kinderen, of geen behandeling meer willen, en andere. Bij de evaluatie van dit ontslag worden deze redenen geïnventariseerd.

Gezien de doelstelling van Overwaal is het van wezenlijk belang dat bij ontslag een *adequate voortgezette hulpverlening* is geregeld die direct aansluit op de klinische behandeling. Na maanden van intensieve behandeling is het uit behandelingstechnisch en menselijk oogpunt noodzakelijk dat de ontslagen patiënt therapeutische begeleiding heeft. Het is de taak van de psychotherapeut dit te regelen. Dat dit niet altijd een eenvoudige taak is, zal een ieder die de ambulante geestelijke gezondheidszorg kent, duidelijk zijn.

De voortgezette hulpverlener en de verwijzer krijgen vervolgens nog een eindverslag toegestuurd, waarin de behandeling op Overwaal en de status quo bij ontslag beschreven staat. Hiermee is dan de bemoeienis met deze patiënt beëindigd, tenzij eventueel een heropname zou worden aangevraagd. In enkele gevallen blijft de psychotherapeut de patiënt ambulant zien, omdat er bijv. (nog) geen hulpverlening gerealiseerd is.

In paragraaf 5.1 werd gesteld dat uit de beschrijving van de werkwijze de evaluatie-momenten zouden kunnen worden afgeleid, alsmede de doelstellingen waaraan het functioneren zou moeten beantwoorden. Uit de gegeven beschrijving, welke is ontleend aan de jaarverslagen over de periode 1974-1985, alsmede op basis van de eigen ervaringen van 9 jaar participatie in de kliniek, kunnen de volgende *doelstellingen* worden afgeleid:

- voldoen de opgenomen patiënten aan de *beoogde populatie* qua problematiek en persoonskenmerken;
- is de *behandelingsduur* in overeenstemming met de A1-norm;
- worden bij ontslag de *doelstellingen bereikt* welke in het behandelingsplan zijn vermeld c.q. waarvoor de patiënt is opgenomen;
- wordt na ontslag de behandeling *tijdig voortgezet* door een ambulante hulpverlener.

Naast deze inhoudelijke doelen kunnen ook enkele *procedures geëvalueerd* worden, die van invloed zijn op het bereiken van de doelstellingen per patiënt c.q. van het instituut als geheel:

- zijn de beoordelingen van de intaker juist;
- worden de doelstellingen door de therapeut voldoende geëxpliciteerd;
- wordt bij ontslag *tijdig voortgezette hulpverlening* geregeld c.q. is de periode tussen ontslag en verdere hulpverlening niet onaanvaardbaar lang.

Tenslotte kunnen nog andere, meer algemeen geldende evaluaties worden uitgevoerd, zoals:

- hoe beoordelen de patiënten de *effectiviteit* van de hulp;
- zijn zij *tevreden* over de behandeling;
- hoe luidt hun *oordeel over de behandelaars* (psycho- en sociotherapeuten);
- wat zijn de *oordelen over de opname zelf*, de kliniek, huisvesting, e.d.

Al deze doelstellingen en overige metingen zullen geëvalueerd worden in het hierna te beschrijven evaluatieproject: *Project Overwaal Patienten/Programma Evaluatie*. Een meta-doelstelling van het project zelf is: *nagaan of het mogelijk is een evaluatiesysteem te ontwerpen dat ook voor andere instellingen bruikbaar kan zijn*. Hierdoor is Overwaal een onderzoeksobject dat een voorbeeldfunctie vervult, een prototype voor programma-evaluatie.

Hoofdstuk 6:

Project Overwaal Programma-evaluatie

*Ik dacht nooit aan de weg,
ik dacht aan het doel.
Elly de Waard (1986).*

6.1 INLEIDING.

In dit hoofdstuk wordt de weg beschreven die *Project Overwaal Programma-Evaluatie (P.O.P.E.)* heeft afgelegd, van zijn start in 1976 tot het eindpunt in 1985. Deze weg is vervolgd met steeds helderder wordende doelen voor ogen, die in meerdere of mindere mate gerealiseerd (blijken te) zijn aan het einde ervan. Aangezien weg en doel met elkaar verband houden, worden deze hier uitgebreid beschreven. Programma-evaluatoren dienen, in tegenstelling tot het citaat boven dit hoofdstuk, beide nauwgezet in het oog te houden. Immers, als doelen niet worden bereikt, is de gevolgde weg wellicht onjuist geweest.

6.2 VOORGESCHIEDENIS.

Van meet af aan werd het doen van evaluatie-onderzoek gezien als een van de doelstellingen van kliniek Overwaal. In het jaarverslag over 1974 was reeds de passage te lezen: "*In de komende jaren hopen wij research op gang te brengen over de behandelingseffecten en over een eenduidig systeem van diagnosestelling*" (Overwaal, 1975). Deze wens om onderzoek te doen, kwam voort uit een dubbele motivatie. Overwaal is een kliniek voor psychotherapie, in het bijzonder *gedragstherapie*. Gedragstherapie wordt gedefinieerd als: "... de methodische toepassing op klinische problemen van experimenteel geverifieerde leerprincipes ..." (Orlemans, 1976, p. 126). De gedragstherapeut zal zich idealiter "... willen bedienen van de empirisch-evaluatieve methode ... hij zal de klinische praktijk zelf willen opvatten als research ..." (p. 129). Samenvattend besluit de auteur dat: "... de gedragstherapeut zich beperkt tot observeerbare gegevens en probeert de therapie zelf als een research-proces te laten verlopen" (p. 130). Door hem wordt als ideaal gezien dat de gedragstherapeut dit moet trachten waar te maken door *zelf onderzoek te doen naar het verloop van een dergelijk proces in de therapie*. Wanneer de betreffende auteur zelf chef de clinique is van een kliniek voor gedragstherapie, is het niet verwonderlijk dat hij tezelfdertijd probeert dit ideaal in de praktijk zoveel mogelijk te benaderen door in deze kliniek te streven naar een zo

objectief en expliciet mogelijke evaluatie van het therapeutisch handelen.

Volgens de grondbeginselen van de gedragstherapie zal in principe iedere gedragstherapeut zijn therapeutisch handelen zoveel mogelijk moeten *expliciteren*, dit in tegenstelling tot een meer intuïtief werkende psychotherapeut. Vervolgens dient hij dan ook dit handelen zo empirisch mogelijk te *evalueren*. Het behoeft geen betoog dat de dagelijkse praktijk vaak veraf staat van deze ideaalstelling en dat evaluatie doorgaans, en dan nog op zijn best, niet verder reikt dan een *subjectieve beoordeling van de bereikte resultaten* welke veelal *impliciet* in het therapieproces liggen opgesloten.

Er zijn veel redenen voor dit verschil tussen ideaal en dagelijkse praktijk. De belangrijkste zullen zijn: het gebrek aan tijd om het behandelingsverloop te expliciteren, een tekort aan kennis en ervaring om objectieve(re) instrumenten te ontwikkelen of de spaarzaam voorhanden zijnde meetprocedures te gebruiken om de resultaten te evalueren. Om dit probleem gedeeltelijk op te lossen, is op Overwaal gekozen voor het uitbesteden van dit evaluatie-onderzoek aan een researchpsycholoog, waarmee bovengenoemde tekorten worden ingevuld. Aangezien er binnen de richtlijnen van het toenmalige COZ geen financiële basis was uit eigen middelen een dergelijke functionaris (part-time) aan te stellen, werd een aanvraag gedaan bij de afdeling psychiatrie van het Sint Radboudziekenhuis te Nijmegen voor detachering van een klinisch-researchpsycholoog. Deze aanvraag werd in 1976 gehonoreerd. Het motief voor deze detachering van de afdeling psychiatrie was o.a. dat zo ervaring kon worden opgedaan met dit type evaluatie-onderzoek dat sterk op de praktijk is gericht. Toepassing van het ontwikkelde en uitgetroefde *evaluatiemodel* op de veel complexere afdeling psychiatrie lag in de bedoeling.

Het *tweede motief* om evaluatie-onderzoek in kliniek Overwaal te verrichten was de behoefte *externe instanties*, met name financiers (het COZ/COTG, e.d.) van het functioneren en de resultaten van de behandeling op Overwaal op de hoogte te kunnen stellen; een behoefte om *verantwoording* af te leggen voor de verleende budgetten. Deze verantwoording werd niet expliciet gevraagd, maar het werd door de staf nuttig geoordeeld reeds op voorhand de behandelingsresultaten aan de financierende instanties kenbaar te maken. Overwaal heeft een relatief hoge stafbezetting en daaruit voortvloeiend een veel hogere verpleegprijs dan gemiddeld in algemene psychiatrische ziekenhuizen. Het periodiek aantonen van de effectiviteit van dit arbeidsintensievere behandelingsmodel leek een verstandige beslissing gelet op het beleid eind jaren zeventig en volgende ten aanzien van de budgetten voor de geestelijke gezondheidszorg, en met name de bezuinigingsdruk hierop. Deze behoefte aan evaluatie zou *extrinsiek* kunnen worden genoemd. Bij de beschrijving van programma-evaluatie zijn bovenstaande twee motieven, het *intrinsieke* en *extrinsieke*, expliciet beschreven (zie hfst. 3).

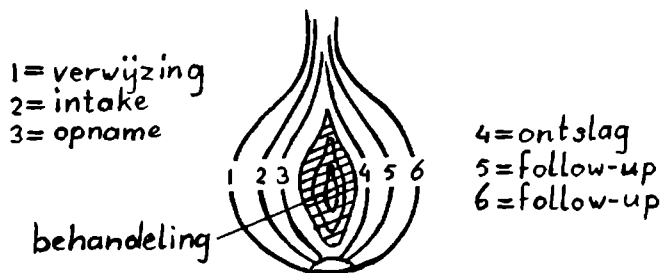
6.3 DOELSTELLING.

Het *einddoel* van het Projekt Overwaal Programma-Evaluatie (P.O.P.E.) is kort gezegd: *het evalueren van verblijf en behandeling van patiënten die in Overwaal zijn opgenomen*. Deze evaluatie kan dienen tot verbetering van opname-indikatie, behandelingsstrategieën, leefklimaat, ontslag-procedure, e.d. Of, met andere woorden, de resultaten van de evaluatie-activiteiten zullen zodanig dienen te zijn, dat ze tot dergelijke verbeteringen concreet aanleiding kunnen geven (Sijben, 1977a).

Hoewel anno 1976 in Nederland het verschijnsel programma-evaluatie nog nauwelijks bekend was, gold hier een typische programma-evaluatieve doelstelling. Ook in de verdere uitwerking blijkt dit: Het project beoogde informatie op te leveren die van nut zou moeten zijn voor de therapeuten. Relevante *feedback* op therapeut- of instituutsniveau zou tevens de motivatie tot participatie aan de evaluatie-activiteiten doen toenemen. Om deze doelstelling te realiseren, werd een *cyclische benadering* voorgesteld. "Dit houdt in dat wij ons nu niet vastleggen over hoe en wat er uiteindelijk geëvalueerd zal worden, maar dat *schils-gewijze* de klinische therapie-processen met evaluatie-instrumenten zullen worden gemeten. Deze heuristische benadering biedt diverse voordelen: de planning van de activiteiten blijft overzichtelijk, de aanmaak en invoer van nieuwe meetinstrumenten is steeds beperkt en hoeft weinig inbreuk te doen op de gang van zaken of een aanslag te plegen op de tijdsinvestering van de medewerkers. Tevens is steeds zicht te krijgen op specifieke invoerproblemen of inhoudelijke onvolkomenheden, welke snel hersteld kunnen worden. Het grote nadeel is dat evaluator en degenen die bij het evaluatie-proces nauw betrokken zijn, de meest essentiële en derhalve ook de frequent gestelde vraag: wat zal je er uiteindelijk mee bereiken, bij de aanvang nog niet kunnen beantwoorden. "Ook gevoelsmatig wordt de harde kern van het evaluatie-object lang ervaren als een steen op de maag, waarvan het nog allerm minst duidelijk is welk laxeer-middel hem te lijf zal kunnen gaan." (Zie Memo 1, Sijben, 1977a, p.3).

6.4 DE OPZET.

De in de vorige paragraaf genoemde *schilsgewijze* benadering werd ook wel aangeduid als het *ui-model* (Sijben, 1977a). Het therapeutisch proces, van verwijzing tot follow-up, kan gezien worden als een *ui*, waarbij iedere *schil* een fase uit het proces voorstelt. De buitenste schillen kunnen doorgaans eenvoudig gepeld worden, echter dichterbij de kern wordt het ontleden van een *ui* veel moeilijker: de schillen kleven sterk aan elkaar vast (zie fig.6.1).



Figuur 6.1 Schilsgewijze evaluatie van een programma volgens het zgn. ui-model.

Het evalueren van verwijzing, intake en opname, evenals het ontslag en een of meerdere follow-up momenten is een activiteit die nauwelijks ingrijpt in het therapeutisch proces zelf, de behandeling. Er wordt weinig meer gevraagd van de behandelaars dan instemming met deze evaluatie-activiteiten en zo nodig hun oordeel over de te gebruiken procedures of instrumenten. Over de *benutting* van de resultaten is hun medewerking echter des te meer nodig, maar daarover later meer. Vooralsnog wordt het evaluatiewerk vóór hen gedaan en krijgen zij te zijner tijd de resultaten gepresenteerd.

Een dergelijke schilsgewijze benadering heeft naast bovenvermeld aspect - aanvankelijk weinig tijd van therapeuten - het voordeel dat de onderzoeker en het instituut geleidelijk in het evaluatieproces groeien: de buitenste schillen worden het gemakkelijkst gepeld, ook onderzoeksmatig, terwijl er geen inbreuk gedaan wordt op organisatorische zaken.

Deze gefaseerde invoering betekent dat eerst de (pre) opname en (post) ontslag momenten in het evaluatie-proces betrokken worden. Zodoende wordt tevens het meest voor de hand liggende meetmodel gehanteerd: *pre-post-design*, terwijl, door enkele meetmomenten in de pre- en postfase te kiezen, er gesproken kan worden van een *time-series-design* (Campbell & Stanley, 1963). De evaluatievraagstelling is primair: *treedt er een verandering op tussen opname en ontslag moment en is deze mogelijke verandering blijvend in de follow-up periode*. Er is dus sprake van een zgn. niet-factorieel "single group design": er worden geen andere groepen onderzocht dan de opgenomen patiënten. In hoeverre opname en behandeling de belangrijkste factor is van mogelijke veranderingen en of deze ook bij niet-opname in dezelfde richting en mate zouden hebben plaatsgevonden, kan niets gezegd worden. Evenmin is duidelijk of een andere behandeling of onderdelen ervan niet tot andere (betere of snellere) resultaten geleid zou hebben. Maar in een programma-evaluatieve opzet is dit ook niet essentieel; waar het om gaat is of alle bemoeienissen en inspanningen *als zodanig* enig effect sorteren. Er wordt impliciet van uitgegaan, dat niet behandelen geen (positief) effect zal opleveren. Uiteindelijk betreft het hier patiënten

die reeds (zeer) lange tijd ernstige problemen hebben, hier soms reeds meer dan 10 jaar voor in behandeling zijn en welke behandelingen (nog) niet tot voldoende resultaat hebben geleid, zodat hun hoop en die van de verwijzer nu op deze (klinische) behandeling is gevestigd. De taak van de hulpverleningsinstelling is met alle hun ten dienste staande c.q. ter beschikking gestelde middelen te trachten het kennelijke dysfunctioneren - in de brede zin van het woord - tot een aanvaardbaar niveau te verminderen. Dan wordt aan de (impliciete) doelstelling van de instelling en dus ook van de voorwaarden scheppende instanties (overheid, financiers) voldaan.

Primair is de opzet van het evaluatieproject gericht op het evalueren van die doelstelling, en is het eerdergenoemde niet-factoriële één-groep-design een toereikend model. Uitbreidingen van dit model kunnen gezocht worden in *herhaalde metingen* in de tijd: naast opname, ontslag en follow-up metingen kunnen ook tijdens de behandeling (deel-)effecten geëvalueerd worden. Daarnaast kan informatie ingewonnen worden bij *meerdere personen*: naast de patiënt ook bij de behandelaar vóór opname, de intaker, de psychotherapeut gedurende de opname en de behandelaar ná ontslag. Te denken valt ook aan personen uit de omgeving van de patiënt, zgn. significante anderen.

Tevens kunnen *meerdere aspecten* van het behandelingsproces en -effect gemeten worden: motieven voor verwijzing, beoordeling van intake criteria, doelstellingen of verwachtingen bij opname en gedurende de behandeling, standaard psychologische vragenlijsten, effect- en satisfactiebeoordeling bij ontslag, beoordeling van ontslag en nazorg procedure, effect en satisfactie beoordeeld bij een of meer follow-up momenten (Zie hoofdstuk 4).

Het is ondoenlijk maar ook onverstandig met een dergelijk uitgebreid informatiepakket een instelling te 'overvallen'. Zoals eerder gezegd dient de weg der geleidelijkheid bewandeld te worden, waarbij zeker de vraag gesteld moet worden welke uitbreiding nog *doelmatig* is. Ook bij evaluatieonderzoek werkt *de wet van de verminderende meeropbrengst* (v.d. Hout, 1984). Echter, om te kunnen beoordelen welke informatie doelmatig is, dient dit als zodanig onderzocht te worden. Dit betekent dat diverse meetinstrumenten en -procedures met elkaar vergeleken moeten worden op die aspecten die de doelmatigheid kunnen beïnvloeden, zoals betrouwbaarheid, bruikbaarheid, validiteit, e.d. Het ligt in de bedoeling dat uit de resultaten van dit project *aanbevelingen* gegeven kunnen worden over de *meest geschikte meetprocedures*.

Deze uitgangspunten, samen met de institutionele en onderzoekstechnische informatiebehoefte, hebben tot een aantal onderzoeksfasen geleid in welke de eerdergenoemde "ui" getracht is te pellen. Hieronder worden deze fasen genoemd en kort beschreven. In de volgende hoofdstukken worden de resultaten van deze evaluatie-activiteiten gepresenteerd.

6.5 DE EVALUATIE-FASEN.

De totale evaluatie-periode bestrijkt *ruim 10 jaar*: van 1974 tot medio 1985, welke in 4 fasen is in te delen:

- 1974 tot 1977: retrospectief follow-up onderzoek, start registratie patiënt gegevens
- 1977 tot 1980: opname, ontslag en follow-up evaluatie
- 1980 tot 1983: vergelijkend onderzoek diverse meetinstrumenten
- 1983 tot 1985: continuering programma-evaluatie.

6.5.1 De periode 1974 tot 1977.

Bij de start van de evaluatie-werkzaamheden medio 1976 werd als eerste activiteit een *follow-up vragenlijst* ontwikkeld. Deze was bestemd voor een *retrospectief onderzoek* bij de patiënten die in de periode 1974, 1975 en 1976 ontslagen waren. In een proefonderzoek onder 50 ex-patiënten werd de vragenlijst en de onderzoeksprocedure getest. Deze werd vervolgens aan ca. 190 ex-patiënten toegestuurd, waarvan 80% werd geretourneerd. De resultaten staan vermeld in hfst. 8.2. Het doel van dit onderzoek was meerledig: therapeuten kennis laten maken met de resultaten van dergelijk evaluatie-onderzoek teneinde aldus hun motivatie en coöperatie te vergroten. Ten tweede werd, ook door de onderzoeker, ervaring opgedaan met dit type onderzoek en de specifieke problemen die zich hierbij kunnen voordoen, zowel inhoudelijk als praktisch.

Er werd ook een zgn. *statistiekformulier* ontwikkeld, dat diende ter registratie van tal van demografische variabelen en opnamegegevens van de patiënten, zoals leeftijd, geslacht, burgerlijke staat, beroep, opleiding, voorafgaande behandeling, verwijzer, opname- en ontslagdatum, psycho- en sociotherapeut en nazorginstantie. De informatie van dit formulier is bedoeld voor het opbouwen van een gegevensbestand van alle opgenomen patiënten. Deze patiëntengegevens werden (uitsluitend op nummer) in een computer opgeslagen. Een dergelijk bestand biedt tal van mogelijkheden om periodiek overzichten te verstrekken van bepaalde gegevens, bijv. ten behoeve van het jaarverslag, alsmede om deze gegevens onderling zowel als met andere informatie in verband te brengen. Deze statistiek-formulieren zijn met terugwerkende kracht vanaf 1974 ingevuld op basis van de statussen. Zoals zovaak, bleek de status een volstrekt ontoereikende informatiebron te zijn. Vanaf 1977 is dit formulier bij *opname* ingevuld en bij *ontslag* gecompleteerd. Aldus werd een gegevensbestand opgebouwd vanaf 1974, en dat verder gedurende de gehele evaluatie-periode is blijven bestaan. In hfst 7.2 worden overzichten van deze statistiekgegevens gepresenteerd.

De *veiligheidsaspecten* van dit en andere opgebouwde gegevensbestanden lijken voldoende gewaarborgd te zijn. De informatie wordt in getalcodes omgezet en verponst. Deze zijn

ingevoerd in de computer van het Universitair Rekencentrum Nijmegen, waar alleen de onderzoeker en (later) de programmeur toegang toe heeft via een variabele "password". Nagaan van de inhoud van een bepaald bestand met een doorgaans cryptische naam is vervolgens alleen mogelijk met kennis van het ponsconcept en de naam-nummer combinatie. Voor derden zijn dit hoegenaamd ondoordringbare barrières, waarachter voor hen weinig zeggende of relevante informatie ligt. Het stelen van statussen uit dossier- of archiefkasten lijkt dan een makkelijker weg om meer inhoudelijke informatie over patiënten te krijgen als men dat zou willen.

Samenvattend: deze *pre-evaluatiefase* wordt gekenmerkt door drie activiteiten: *ontwikkelen van een registratieformulier* voor patiëntengegevens en het met terugwerkende kracht invullen hiervan, het verrichten van een *retrospectief follow-up onderzoek* en ten derde het *scheppen van condities* voor het verder verrichten van evaluatie-onderzoek zoals het motiveren van medewerkers, het verwerven van een positie van een onderzoeker "van buiten die erbij hoort" door bekend te raken met alle interne behandelingsprocedures, het bijwonen van alle intake- en teamvergaderingen, het zelf verrichten - op bescheiden schaal - van curatieve werkzaamheden. Daarnaast dienden ook organisatorische condities geschapen te worden, zoals het verwerven van administratieve ondersteuning, e.d. Op Overwaal waren nauwelijks of geen faciliteiten ten behoeve van onderzoek aanwezig; alle middelen dienden door de K.U. Nijmegen verstrekt te worden, zoals reprografie, computerfaciliteiten, postverzending, etc. Het leren hiervan optimaal gebruik te maken was een vereiste voorwaarde voor verdergaan- de en arbeidsintensievere onderzoeksactiviteiten.

Onderzoekers in de sociale wetenschappen wordt dikwijls terecht verweten zich in een *ivoren toren* te plaatsen. Hier was het heel nadrukkelijk de bedoeling onderzoek in en ten behoeve van de praktijk te verrichten. De eerste werkkamer, een isoleercel, leek niet bij dit beeld passen; het beantwoordde niet aan de bedoeling *onderzoek te assimileren en integreren in de curatieve zorg*. De volgende fasen moeten leren of dit in de loop der jaren is gelukt. - Voor een betere werkkamer was overigens snel gezorgd.-

6.5.2 De periode 1977 tot 1981.

De ervaringen met het statistiekformulier en de ontwikkelde follow-up vragenlijst waren zeer bevredigend. Het statistiekformulier werd m.i.v. 1977 reeds bij opname ingevuld, aanvankelijk door de secretaresse en/of intaker. Deze procedure werkte niet, zodat de onderzoeker dit samen met de patiënt deed. Hij was dan tevens in de gelegenheid iedere patiënt persoonlijk in te lichten over de evaluatiewerkzaamheden. In een ca. 15 minuten durend gesprek met de nieuw opgenomen patiënt(en) werd hem (hen) uitgelegd wat de positie van de onderzoeker was (medewerker K.U. afdeling psy-

chiatrie), het motief van dit onderzoek (intrinsiek en extrinsiek), de aard van de informatie die aan hen gevraagd zou worden (evaluatie van hun behandeling), wat met deze gegevens gebeurde (periodieke overzichten voor interne en externe belanghebbenden) en het vertrouwelijke karakter van dit onderzoek (computeropslag op nummer).

Het proefonderzoek met de follow-up lijst was van dien aard dat deze lijst werd aangepast, zodat deze reeds bij ontslag kon worden ingevuld. Met ingang van 16 november 1976 werd deze *zgn.ontslagvragenlijst* aan de patiënten ca. één week vóór ontslag uitgereikt. Vervolgens werd hen na 3, 6 en 12 maanden nog een *follow-up vragenlijst* toegezonden.

Sinds 1977 werd door Overwaal deelgenomen aan het patiënten registratie systeem (P.I.G.G) van het Ministerie van Volksgezondheid (Leidschendam). Hiertoe dienden ook centraal diagnoses verzameld te worden. Daarom werd er een *zgn. diagnoseformulier* ingevoerd, waarop de psychotherapeut een of meer diagnoses (I.C.D.-9, 1978) kon invullen, alsmede de mate waarin de patiënt aan deze diagnose(s) leed (ernstscore). In de loop van 1977 werd vervolgens ook een *ontslagvragenlijst voor de psychotherapeut* ingevoerd en *drie follow-up vragenlijsten voor de voortgezette hulpverleners*. Voor de verwijzers werd een *zgn. aanvraagformulier* ontworpen om aard van verwijzer en motieven voor deze verwijzing te registreren. Uitbreiding van deze lijst met een opgave van problematiek en/of diagnose en andere informatie, zo mogelijk in (semi-)gestandaardiseerde vorm, stuitte op weerstand bij de verwijzers. Zij gaven er in meerderheid de voorkeur aan een (uitgebreide) aanmeldingsbrief te schrijven, waarin ze naar eigen behoefte en inzicht informatie over de aangemelde patiënt zouden kunnen vermelden.

Vanaf 1978 bestond aldus een opname-, ontslag- en follow-up-evaluatieschema, zoals dit in tabel 6.1 is weergegeven. Dit schema is gedurende de alle evaluatieperiodes blijven bestaan; er werden in de volgende fase wel inhoudelijke veranderingen in de lijsten en uitbreidingen van de informatie-verzameling aangebracht.

Hieronder wordt een korte opsomming gegeven van de inhoud van de in tabel 6.1 vermelde vragenlijsten:

- *Aanvraagformulier*: naam aanvrager, discipline en instelling waar deze werkzaam is, mate waarin hij op de hoogte is van de problematiek, motieven van verwijzing en of hij eventueel bereid is ook de voortgezette hulpverlening na ontslag op zich te nemen.
- *Statistiekformulier*: personalia en overige demografische gegevens van de patiënt, duur van de klachten en eerdere behandelingen, aard van deze behandelingen, mate van dysfunctioneren, opname- en ontslagdata.
- *Diagnoseformulier*: maximaal 5 psychiatrische diagnose codes, in volgorde van belangrijkheid, ernstscore per diagnose (5-puntsschaal), maximaal 3 somatische diagnoses, is suicide(-dreiging) mede reden tot opname.

Tabel 6.1 Overzicht informatie/evaluatie-systeem
zoals dit sinds 1978 bestond.

LIJST	IN GEVULD DOOR	MOMENT
aanvraagformulier	verwijzer	verwijzing
statistiekformulier	onderzoeker + pat.	opname
diagnoseformulier	psychotherapeut	1e maand
ontslagvragenlijst	patiënt	ontslag
ontslagformulier	psychother.	ontslag
1e follow-up	patiënt	3 mnd. na ontslag
1e follow-up nazorg	voortg. hulp	3 mnd. na ontslag
2e follow-up patiënt	patiënt	6 mnd. na ontslag
2e follow-up nazorg	voortg. hulp	6 mnd. na ontslag
3e follow-up patiënt	patiënt	12 mnd. na ontslag
3e follow-up nazorg	voortg. hulp	12 mnd. na ontslag

- *Ontslagvragenlijst patiënt*: beoordeling ernst opnameproblemen, reden verwijzing, effect behandeling, beoordeling behandelingsonderdelen, verblijf en ontslag van Overwaal.
- *Therapeut ontslag vragenlijst*: diagnoses bij ontslag, ernstscore, reden ontslag, beoordeling ontslagprocedure, regeling voortgezette hulpverlening, beoordeling ontslagresultaat.
- *Follow-up vragenlijsten patiënt*: beoordeling status quo, verschil tussen ontslag en follow-up moment, bron van inkomsten, mate van disfunctioneren, beoordeling voortgezette hulpverlening, behoefte aan (andere) hulpverlening, beoordeling opname op Overwaal achteraf (alleen 1e follow-up).
- *Follow-up vragenlijst nazorg*: beoordeling wijze waarop nazorg geregeld is door Overwaal (alleen 1e follow-up), duur, frequentie van de behandeling, vorm en inhoud ervan, beoordeling status quo en vorderingen sinds ontslag, oordeel behandeling op Overwaal (alleen 1e follow-up).

Deze set formulieren en vragenlijsten werden tot 1981 in deze vorm gebruikt. Zij leverden een permanente stroom van registratieve en evaluatieve informatie die op diverse niveau's werd teruggekoppeld. De resultaten staan vermeld in hoofdstuk 7 en 8.3.

6.5.3 De periode 1980 tot 1983.

Na enige jaren met deze lijsten gewerkt te hebben, kwamen ook de manco's aan het licht: *de evaluatie van de doelstellingen was te globaal en alleen achteraf*. Ook enkele antwoordcategorieën waren te weinig over de schaal verspreid of vertoonden andere tekstuele onvolkomendheden. Sommige vragen

waren te redundant, terwijl andere informatie ontbrak. Daarom werden in 1980 alle vragenlijsten herzien, waarbij deze ook beter op elkaar werden afgesteld zodat uniforme longitudinale informatie verkregen werd.

Een andere vernieuwende impuls werd gegeven door de ideeën over *programma-evaluatie*. Na een literatuurstudie werd in jan./febr.1979 een studiereis naar de U.S.A. gemaakt om kennis te nemen van deze evaluatievorm in de praktijk. Er werden diverse "mental health" instellingen bezocht, zowel klinische als ambulante, in Minneapolis, Denver, Los Angeles en San Francisco die reeds enige jaren ervaring hadden met programma-evaluatie (Sijben, 1979a).

De ervaringen, opgedaan in deze reis, mondden uit in een *onderzoeksplan* voor Overwaal, waarin enerzijds het bestaande evaluatiesysteem werd verbeterd om *permanent feedback* te kunnen leveren over kwaliteit en effectiviteit van de hulpverlening. Anderzijds dienden diverse meetinstrumenten met elkaar *vergeleken* te worden om na te gaan welke de meest *bruikbare* zijn. Omdat een dergelijk project niet met de beschikbare menskracht kon worden uitgevoerd, werd een *subsidie-aanvraag* (Sijben, 1979b) ingediend bij het *Praeventiefonds* voor ondersteunend personeel: administratief (0,8) en programmeur (0,2) en voor bijkomende onkosten. Deze subsidie werd met ingang van 1 jan. 1980 aanvankelijk voor 4 jaar toegekend en bedroeg in totaal f.230.000,-.

Tevens werd bij de Katholieke Universiteit Nijmegen computerapparatuur aan gevraagd voor lokale opslag van gegevens. Deze werd eveneens gehonoreerd. Aldus kon met ingang van 1980 de meest intensieve evaluatiefase starten.

Tot nu toe lagen de meetmomenten met name aan het *begin* en aan het *einde van de behandeling* en bevatten inhoudelijk vooral algemene patiëntgegevens en beoordeling van de satisfactie en globale veranderingen van de patiënt. De door het Praeventiefonds verleende subsidie maakte het mogelijk de arbeidsintensievere evaluatie *tijdens* het behandelingsproces in het onderzoek te betrekken. Ook de inhoud van de voor- en nametingen diende veranderd en uitgebreid te worden.

Zoals hierboven is vermeld, werden, op basis van de ervaringen, de sinds 1977 gebruikte vragenlijsten grondig herzien: naast de *satisfactievragen* en *beoordelingen van de procedures* e.d. werden o.a. bij opname *doelstellingen* geïnventariseerd en bij ontslag geëvalueerd. Een registratieformulier werd toegevoegd: een intakeformulier waarop de intaker kon aangeven in hoeverre de opnamecriteria van toepassing zijn geweest. Deze werden door de psychotherapeut bij ontslag geëvalueerd.

Tevens werden *standaard psychologische vragenlijsten* ter meting van angst, depressie, dwang, klachten, etc. ingevoerd. Deze instrumenten, die vooral worden gebruikt in vergelijkend psychotherapie-effect onderzoek, konden nu met de satisfactie- en doelmetingen worden vergeleken. Aldus ontstond een "*multitrait-multimethod design*" (Campbell & Fiske, 1959), hetgeen de gewenste vergelijkende informatie diende te verschaffen om de meest bruikbare evaluatie-instrumenten te

kunnen aanwijzen. Hieronder volgt een opsomming en motivering van de gebruikte meetprocedures welke naast de in paragraaf 6.4.2 genoemde vragenlijsten, werden ingevoerd.

6.5.3.1 Psychologische vragenlijsten.

In hfst. 4 is uitgebreid aandacht besteed aan het meten van psychotherapie-effecten. De met name in Nederland vaak gebruikte instrumenten zijn *psychologische vragenlijsten*. Gelet op de problematiek van de op Overwaal opgenomen patiënten (zie ook par. 7.3, diagnoses) lag het voor de hand te kiezen voor lijsten die pretenderen: dwang, globale angst, depressie, fobieën, lichamelijke klachten, sociale angst resp. assertiviteit, persoonlijkheidskenmerken en welbevinden te meten. Anno 1980 waren daarvoor nog slechts een beperkt aantal instrumenten beschikbaar. Keuze van instrumenten voor het meten van psychotherapie-effecten lijkt vaak arbitrair, waarbij *beschikbaarheid* doorgaans een belangrijkere rol lijkt te spelen dan andere eigenschappen, zoals te meten concepten, psychometrische kwaliteiten, e.d. De keuze voor deze vragenlijsten in dit project werd voornamelijk ook bepaald door het feit dat deze in diverse therapie-effectstudies werden gehanteerd, waardoor ze enige reputatie hadden verworven op dit gebied. Nagegaan kan nu worden of de binnen het project ontwikkelde instrumenten hiermee vergelijkbare informatie opleveren. Hieronder volgt een beschrijving van de dimensies die door deze standaard psychologische vragenlijsten gemeten worden.

- *Nederlandse Persoonlijkheids Vragenlijst (N.P.V.)* (Luteijn, 1975). Deze vragenlijst meet een 7-tal persoonlijkheidskenmerken:

1. Inadequatieschaal (IN): deze schaal bevat vage lichamelijke klachten, een gedrukte stemming, vage angsten en insufficiëntiegevoelens.
2. Sociale Inadequatieschaal (SI): de items van deze schaal verwijzen naar het vermijden van of het zich ongelukkig voelen in sociale contacten.
3. Rigiditeitschaal (RG): de items uit deze schaal hebben te maken met het volgens plan willen laten verlopen van gebeurtenissen en met vaste gewoontes en principes.
4. Verongelijktheidschaal (VE): in de items van deze schaal staat kritiek op en wantrouwen van andere mensen centraal.
5. Zelfgenoegzaamheidschaal (ZE): in de items van deze schaal komt een gevoel van tevredenheid over zichzelf en zich niet voor een ander en diens problemen willen interesseren naar voren.
6. Dominantieschaal (DO): de items van deze schaal hebben te maken met zelfvertrouwen, initiatief willen nemen en leiding willen geven aan anderen.
7. Zelfwaarderingsschaal (ZW): de items uit deze schaal verwijzen naar een positieve instelling t.o.v. werken, veranderbaar en goed aangepast zijn.

- *Fear Survey Schedule (F.S.S.)* (Wolpe & Lang, 1964), ook wel bekend als Vreeslijst. Deze vragenlijst inventariseert welke situaties in welke mate angst oproepen. Factoranalytisch onderzoek (Arrindell, 1980) heeft het bestaan van 5 dimensies aangetoond: sociale angst, straatvrees, angst voor ziekte, dood, e.d., angst voor seksuele of agressieve uitingen, angst voor onschuldige dieren. Een niet door Arrindell empirisch geverifieerde 6e dimensie is angst voor plotselinge of harde geluiden (Grossberg & Wilson, 1965).
- *Manifest Anxiety Scale (M.A.S.)* (Taylor, 1953; von Bonninghausen, z.j.) meet slechts een factor: globale angst, welke gedefinieerd wordt als: overt or manifest symptoms of the state of anxiety or emotionality (Taylor, p. 285). De items zelf zijn afkomstig uit de M.M.P.I.
- *Sociale Angst Schaal* (Willems e.a., 1973). De schaal meet de angst die iemand heeft voor de negatieve waardering door anderen voor wat betreft zijn sociale vaardigheid (presentatie) en zijn fysieke verschijning (uiterlijk). De lijst van 24 items pretendeert sociale angst als dispositie te meten.
- *Assertiviteitslijst* (Brinkman, 1977). Deze in Nederland ontwikkelde lijst bevat 6 dimensies van assertief gedrag:
 1. Beëindigen van onaangename situaties.
 2. Angst voor aversieve sociale stimuli.
 3. Weigeren.
 4. Uiten van emoties.
 5. Streven naar/acceptatie van positieve sociale stimuli.
 6. Informatie vragen, tegenspreken, discussiëren.

T. De totaalscore.

Deze indeling is gemaakt op basis van de inhoud van de items. Factoranalytische resultaten waren er in 1980 nog nauwelijks. Later bleek evenwel dat deze 6 dimensies niet in een factorstructuur werden weerspiegeld (Prudon, interne mededeling)
- *Depressieschaal* (Zung, 1965). Deze vragenlijst meet de mate van depressie middels het kwantificeren van de symptomen ervan, "... using the diagnostic criteria of the presence of a pervasive depressed affect, and its physiological and psychological concomitants as test items" (Zung, 1965, p. 70). De lijst wordt ook in Nederland veel gebruikt met goede resultaten (Dijkstra, 1974).
- *Inventarisatie Dagelijkse Bezigheden (I.D.B.)* (Kraaimaat & Van Dam-Baggen, 1976). Deze lijst, gebaseerd op de Leyton Obsessional Card Inventory (Cooper, 1970) is een vragenlijst die obsessieve/compulsieve symptomen registreert. Uit onderzoek blijkt dat deze vragenlijst betrouwbaar is en goed discrimineert tussen normalen en obsessieve/compulsieve patiënten.
- *Innerlijk Welbevinden Schaal* (Hermans, 1973; Visser, 1978). Deze vragenlijst bevat twee dimensies: positieve en negatieve gevoelens. Bij nadere bestudering bleek de in dit onderzoek gebruikte vragenlijst af te wijken van de door Hermans ontwikkelde schaal. De verschillen zijn:

I.W.-Hermans: 36 items waarvan 17 positief en 19 negatief. Onze versie: 37 items waarvan 20 positief en 17 negatief. Er blijken in onze versie 9 vragen voor te komen die niet in de oorspronkelijke versie staan en omgekeerd worden 8 vragen uit de Hermans-versie gemist. Navraag leverde geen informatie op over de herkomst van de gebruikte versie en de motieven voor verandering. Gezien de grote overlap met de Hermans-versie en de inhoudelijke overeenkomstigheid van de verschillende items wordt deze schaal toch verder bij de analyses betrokken, waarbij de resultaten voorzichtig geïnterpreteerd zullen worden en de door Hermans gevonden gemiddelde waarden en correlaties dus niet bruikbaar zijn. Het is natuurlijk onjuist een bestaande vragenlijst eigenhandig te veranderen, terwijl dit niet in de naamgeving of anderszins wordt kenbaar gemaakt zonder overleg met de auteur. Dit brengt anderen op een dwaalspoor.

- *Eysinck Personality Inventory* (Eysinck, 1964). Deze 12 items tellende lijst meet 2 dimensies: neuroticisme en extraversie. Onderzoek van Eysinck toont aan, dat deze vragenlijst betrouwbaar is terwijl ook de validiteit zeer positief beoordeeld wordt (Buros, 1972). Negen items komen ook in de A.B.V. (Wilde, 1963) voor. Nederlandse gegevens zijn niet bekend. Deze lijst is later in het pakket toegevoegd. Een vergelijkend onderzoek met de A.B.V. wordt nog verricht.

Aldus een opsomming van de gebruikte standaardvragenlijsten. Deze vragenlijsten, in totaal 503 items, werden afgenomen bij opname, ontslag en 6 maanden follow-up. In tabel 6.2 worden deze nog kort weergegeven.

Tabel 6.2 Standaard psychologische vragenlijsten

Naam	N	Variabele
Ned. Persoonlijkheids Vragenlijst (N.P.V.)	133	Diverse Persoons kenmerken
Fear Survey Schedule (F.S.S.)	76	Specifieke fobieën
Manifest Anxiety Scale (M.A.S.)	52	Globale angst
Sociale Angst-Willems (S.A.)	24	Sociale angst
Assertiviteitslijst-Brinkman	60	Diverse aspecten van assertiviteit
Depressie-Zung (Z.D.S.)	20	Depressie
Inventarisatie Dagelijkse Bezigheden (I.D.B.)	32	Dwang
Innerlijk Welbevinden (I.W.)	37	Gevoelens van pos. en neg. welbevinden
Klachtenlijst-Brinkman	57	Lichamelijke klachten
Eysinck Personality Inventory	12	Neuroticisme en Extraversie

Het was een van de hypothesen bij de start van het project dat de hiervoor beschreven en veel gebruikte standaard psychologische vragenlijsten niet (alleen) toereikend zijn om behandelingseffect te meten, en wel om meerdere redenen: deze lijsten geven vaak slechts *globale informatie*, zoals mate van angst of neuroticisme, welke niet altijd correspondeert met de problematiek van de patiënt en/of het behandelingsdoel. Veranderingen in probleemgedrag - het primaire doel van iedere behandeling - worden dan niet of nauwelijks weerspiegeld in veranderde uitslagen op deze vragenlijsten. Een ander probleem is de evaluatie van deze veranderingen, die eerder *relatief* of *normatief* (i.e. afhankelijk van de groepsgegevens) dan absoluut of *criterium georiënteerd* worden getoetst. Dit betekent dat aan een verschil tussen begin- en eindscore betekenis wordt toegekend op basis van groepsgemiddelde en standaarddeviatie, terwijl dit verschil weinig inhoudelijke betekenis kan hebben (cf. klinische relevantie versus statistische significantie, Meehl, 1954). Aan dit probleem is in Hfst. 4 uitvoerig aandacht besteed.

Doelevaluaties komen gedeeltelijk aan deze bezwaren tegemoet: de verzamelde informatie is patiënt-gebonden, de problemen en/of behandelingsdoelen worden aan het begin van de behandeling gespecificeerd alsmede het gewenste behandelingseffect. Bij de beëindiging van de therapie wordt vervolgens nagegaan in hoeverre dit effect is gerealiseerd. Er is nu sprake van een criterium georiënteerde evaluatie. Een probleem dat zich hierbij voordoet is: wie formuleert de doelstellingen? Uit het geruchtmakende onderzoek van Hageman-Smit (1976): *De Client en zijn hulpverlener, een paar apart*, bleek dat er grote verschillen kunnen bestaan tussen de doelen die de cliënt en zijn hulpverlener formuleerden. In dit project is daarom gekozen voor drie informanten: de patiënt alleen, de therapeut al dan niet in overleg met de patiënt en de onderzoeker samen met de patiënt. Dit leidt tot drie doelevaluaties, waarvan achteraf nagegaan kan worden in hoeverre deze inhoudelijk verschillend zijn en of deze tot verschillende evaluatie-scores aanleiding geven. Van deze drie doelevaluatiemethoden wordt hier kort de procedure geschetst:

- Individuele Probleem Evaluaties (I.P.E.).

Bij opname van de patiënt vulde deze o.a. de zgn. *Opnamevragenlijst* in. Deze bevatte de vraag: Wilt u hier een, twee of hooguit drie klachten/problemen kort beschrijven en per probleem aangeven hoe erg u eronder lijdt? Hierna volgde ruimte voor het beschrijven van deze problemen met per probleem de vraag: Hoe erg lijdt u aan dit probleem? met een vijf-puntsschaal: van 1 = heel erg tot 5 = niet. Bij het uitdelen van deze lijst werden de patiënten nadere aanwijzingen gegeven hoe ze deze problemen konden beschrijven: geen grote verhalen, maar ook niet volstaan met een of twee woorden, zoals klachten of angsten. Er werden dan enkele voor-

beelden gegeven. Daarnaast werd voorgesteld bij moeilijkheden of onduidelijkheden bij het invullen van deze of andere lijsten hiermee terug te komen of dit bij het inleveren te bespreken. Tevens werd gevraagd per probleem de *te verwachten verandering* aan te geven op een vijf-puntsschaal: van geheel verbeterd tot niet verbeterd.

Deze I.P.E. werden vervolgens gecodeerd volgens een eigen ontwikkelde rubriekenlijst (zie verder) en op een apart vel overgetypt. Bij de ontslagvragenlijst werd dit vel toegevoegd. In de lijst werd hiernaar verwezen en vervolgens per probleem gevraagd of dit op Overwaal *behandeld* was en of dit probleem *verbeterd* was en in welke mate (zes-puntsschaal). Tevens had de patiënt de mogelijkheid twee nieuwe problemen te noteren die gedurende de behandeling bekend waren geworden en vervolgens ook deze te evalueren. Bij de drie follow-up momenten werden deze, maximaal 5, problemen wederom geëvalueerd.

- Individuele Doel Evaluatie therapeut/patiënt (I.D.E.).

Deze en de volgende (I.D.E.E., zie beneden) *doelevaluatieprocedures* zijn geïnspireerd op de Goal Attainment Scaling (G.A.S.) Voor een beschrijving van deze methode wordt verwezen naar hoofdstuk 4.3.2. Reeds in 1978 is op Overwaal een proef genomen met deze G.A.S., maar deze is mislukt: de methode is voor de praktijk te *bewerkelijk*, de therapeut moet teveel informatie genereren, waarbij hij een exactheid moet demonstreren die hem volkomen vreemd is op het terrein van de beoogde behandelingseffecten. De na te streven veranderingen zijn doorgaans slechts impliciet aanwezig.

Deze I.D.E.-procedure is daarom in enkele opzichten vereenvoudigd in vergelijking met de G.A.S. De therapeut dient binnen de eerste weken van de opname op een formulier aan te geven welke doelstelling hij voor deze patiënt expliciteert. Vervolgens geeft hij aan: hoe hij vorderingen kan meten m.a.w. de operationalisatie(s) ervan - dit kunnen er meerdere zijn -, door wie de meting verricht wordt en over welke periode. Hij vermeldt dan de *beginscore* met de datum waarop deze is gemeten en de zgn. *streefscore*, eveneens gedateerd. Dit formulier wordt vervolgens ingeleverd en de doelstellingen worden van codes voorzien. Deze doelen moeten iedere maand geëvalueerd worden. Deze evaluatie betreft een globale beoordeling van de vooruitgang van de doelstelling richting het geformuleerde doel in percentages (100%: doel is bereikt), zowel door de therapeut als de patiënt ingeschat. Tevens wordt de motivatie van de patiënt aangegeven op een vijf-puntsschaal. Daarna volgt een evaluatie van de *geoperationaliseerde metingen*, waarbij per meting de nieuwe score bepaald en genoteerd moet worden.

Aanvankelijk werden deze I.D.E.'s overgetypt en maandelijks geretourneerd, zodat de therapeut de nieuwe scores kon toevoegen, maar deze procedure was te *bewerkelijk* en te arbeidsintensief (Sijben, 1981). Door ir. N. Cox, U.R.C. Nijmegen, is een programma ontwikkeld, waardoor deze doelstel-

lingen in een computer kunnen worden ingevoerd en waaraan later scores worden toegevoegd of mutaties kunnen worden aangebracht. Hierna wordt een nieuw formulier afgedrukt dat door de therapeut bij de volgende evaluatie kan worden ingevuld. Deze metingen worden bij ontslag voor het laatst genoteerd. Nagegaan kan nu worden in hoeverre de doelen bereikt zijn i.c. of de streefscore overeenkomt met de werkelijk behaalde score.

- Individuele Doel Evaluatie Extern (I.D.E.E.).

Door de onderzoeker worden in de eerste twee weken van de opname in een ca. 1 uur durend *interview* eveneens doelstellingen geïnventariseerd en in meetbare termen vastgelegd. In dit gesprek worden eerst de probleemgebieden geïnventariseerd (Wat zijn de klachten of problemen waarvoor u naar Overwaal bent gekomen). Vervolgens wordt per probleem doorgevraagd om inzicht te krijgen in aard, intensiteit, frequentie e.d. ervan op het moment van opname. Vervolgens wordt gevraagd wat de patiënt verwacht van zijn opname op Overwaal (Zoals u weet volgt meestal na uw opname nog voortgezette hulpverlening, dus uw problemen zullen niet over zijn bij ontslag, maar hopelijk zozeer verbeterd dat u weer naar huis kunt om daar verder te gaan). Als u over enkele maanden weer ontslagen wordt, hoe zouden uw problemen dan verbeterd moeten zijn). Deze verwachtingen worden eveneens in operationele termen vastgelegd.

Deze doelen worden bij ontslag en 6 maanden hierna geëvalueerd. In een gesprek worden de doelen en verwachtingen ter sprake gebracht en wordt nagegaan hoe de status quo is. Voor het follow-up gesprek wordt de patiënt uitgenodigd op Overwaal te komen (reiskosten werden vergoed). Incidenteel werden ze thuis bezocht of vond het gesprek bij de onderzoeker thuis plaats als er motieven waren om niet op Overwaal te komen (niet willen/durven, de moeite niet willen nemen, e.d.). Bij het ontslag- en follow-up gesprek worden ook de uitslagen van de standaardvragenlijsten besproken.

Alle doelstellingen (I.P.E., I.D.E. en I.D.E.E.) werden gecodeerd. Hiertoe werd een *rubriekenlijst* ontwikkeld op basis van bestaande diagnostische classificaties (I.C.D.-9, 1978; DSM-III, 1980) en de classificatie van Schutz (1981) aangevuld met rubrieken naar aanleiding van concrete doelstellingen. Uiteindelijk bevatte deze lijst 20 hoofdrubrieken met elk meerdere onderverdelingen. De hoofdrubrieken waren: Lichamelijke klachten, Fobieën, Dwang, (vage) Angsten, Sociaal dysfunctioneren, Emotioneel dysfunctioneren, Rouwproblemen, Relatieproblemen, Verslaving, Depressie, Tics, Problemen met lichaamsuitscheiding, Anorexia Nervosa, Faalangst, Sexueel dysfunctioneren, Dysfunctioneren in het dagelijks leven op basis van problemen, Problemen door gedachten of cognities (m.b.t. gebeurtenissen, e.d.), Overige problemen en Problemen m.b.t. het therapie-programma of -proces. Een niet in te delen rubriek completeerde de lijst. (Zie ook: Classens, 1983). Deze viercijferige code werd aan de gegevens

toegevoegd om inhoudelijke vergelijkingen tussen de diverse doelmetingen mogelijk te maken alsmede per hoofdrubriek evaluaties te kunnen geven.

- Medicijngebruik.

Naast alle hierboven beschreven - hoofdzakelijk evaluatieve - gegevens werd ook per patiënt het *medicijnverbruik* geregistreerd. Dat hield in dat door de artsen op een lijst werd genoteerd aan wie hij welke medicatie met welke dosering voorgeschreven had. Deze lijsten werden per maand verzameld, gecontroleerd aan de hand van de nota van de apotheek en op patiëntnummer ingevoerd. Hiertoe waren alle medicijnen gecodeerd volgens een rubricering opgesteld in overleg met Dr. M. Hoes (medewerker afd. Psychiatrie).

Aldus bestond er een heel *netwerk van registratie- en evaluatie-instrumenten* dat inhoudelijk en longitudinaal aan elkaar gerelateerd was of kon worden. In tabel 6.3 staat een overzicht van alle metingen, het tijdstip van afname en de bron/invuller.

Dit hele stelsel moet de meest relevante informatie opleveren over het curatieve functioneren van deze klinische setting. De resultaten hiervan staan vermeld in Hfst. 8 en 9.

6.5.4 De periode 1983 - 1985.

Na drie jaar op de hiervoor beschreven intensieve wijze het curatieve proces geëvalueerd te hebben, was het om meerdere redenen gewenst deze informatieverzameling te beëindigen c.q. te verminderen. Aanvankelijk werd voorgesteld het afnemen van de standaardvragenlijsten af te schaffen, maar enkele psychotherapeuten stelden hier evenwel prijs op. Uit de eerste ervaringen was echter gebleken, dat niet alle lijsten even goed bruikbaar waren en dat inmiddels andere lijsten bekend waren.

Zo was in een onderzoek (Classens en Gommans, 1983) de *Symptoms Checklist, SCL-90* (Derogatis e.a., 1973, 1977a, 1977b) in de door Arrindell & Ettema (1981) vertaalde versie goed bruikbaar gebleken als klachtenlijst. Daarnaast was behoefte om de *Eysinck Personality Inventory* te relateren aan de *Amsterdamse Biografische Vragenlijst*, zodat deze beide in 1983 werden afgenomen. Daarnaast kon de Zung Depressie Schaal vervallen, omdat de SCL-90 een depressieschaal bevat. Eveneens verviel de I.W., omdat deze lijst weinig bruikbaar bleek. De overige vragenlijsten bleven gehandhaafd. Dit leidde tot de lijsten met hun dimensies zoals in tabel 6.4 vermeld staat.

Dit pakket vragenlijsten, in totaal 393 vragen, een vermindering met 22%, is nog een rijkelijke hoeveelheid. Uit vergelijkingen van de uitslagen zou kunnen blijken dat de E.P.I. in grote mate overeenstemt met de A.B.V. en dat subschalen van de SCL-90 overeenkomen met de scores op bijv. de

Tabel 6.3 Totale meetprocedure P.O.P.E. (1980-1984).

<u>Aanmelding</u> & <u>Opname:</u>	<u>Opname-periode</u>				<u>Ontslag</u>	<u>Follow-up</u>		
	1 ^e mnd.	2 ^e mnd.	3 ^e mnd.	laatste maand		na 3 mnd.	na 6 mnd.	na 12 mnd.
- Aanvraagform.								
- Statistiek- formulier								
- Opnamelijst					Ontslaglijst	F.U.P.	F.U.P.	F.U.P.
- P-kaart					P-kaart			
- Diagnoseform.					Diag.ontsl.			
- I.D.E.E.					I.D.E.E.M.		I.D.E.E.M.	
- I.D.E.	I.D.E.M.	I.D.E.M.	I.D.E.M.	I.D.E.M.				
- Psychol.lijsten					Psychol. lijsten		Psychol. lijsten	
					Ther.ontslag lijst	F.U.N.	F.U.N.	F.U.N.

Verklaring: I.D.E. = Individuele Doel Evaluatie (therapeut)

I.D.E.E. = Individuele Doel Evaluatie Externebeoordelaar

I.D.E.(E.)M. = Individuele Doel Evaluatie (Externe) Meting

P-kaart = Patiëntenregistratie intramurale geestelijkegezondheidszorg

F.U.P. = Follow-Up Patiënt

F.U.N. = Follow-Up Nazorg

Tabel 6.4 Vragenlijsten afgenomen sinds 1983

VRAGENLIJSTEN	N	MEETDIMENSIES
A'damse Biografische Vragenlijst (ABV)	107	Neuroticisme, Neuro-Somatische klachten, Extraversie, Testattitude
Symptom Check List (SCL)	90	Agorafobie, Angst, Depressie, Somatisering, Slaap, Interpersoonlijke sensitiviteit, Obsessie, Hostiliteit, Totalscore
Fear Survey Schedule	76	Diverse fobische angsten
Sociale Angstschaal	24	Sociale angst
Manifest Anxiety Schaal	52	Globale angst
Inventarisatie dag. bezigheden	32	Dwang
Eysinck Personality Inventory	12	Neuroticisme, Extraversie

F.S.S., S.A., M.A.S. en I.D.B. In dat geval zou het vragenlijstenpakket teruggebracht kunnen worden tot 102 vragen (SCL-90 en E.P.I.). Dit alles bij de veronderstelling dat *dergelijke informatie aan een behoefte voldoet en/of diagnostische of evaluatieve waarde heeft*. Het verifiëren van deze veronderstellingen komt in Hfst. 10 verder aan bod. De doel-evaluaties waren eveneens onderwerp van discussie:

- *De Individuele Probleem Evaluaties* (patiënt) bleven bestaan, aangezien deze waren opgenomen in de evaluatie-vragenlijsten (opname-, ontslag- en follow-up lijsten) en reeds zeer nuttig bleken te zijn.

- *De Individuele Doel Evaluaties* (therapeut) waren slechts door enkele therapeuten systematisch ingevuld. Toen werd voorgesteld deze af te schaffen, kwam van deze therapeuten echter bezwaar: zij wilden hier zelf graag mee doorgaan. Op facultatieve basis is dit dan ook gebeurd en werd ook de computerondersteuning gecontinueerd. Maar dit bleek toch snel te verwateren, en in de loop van 1984 stopte deze vorm van evaluatie.

- *De Individuele Doel Evaluatie Extern (onderzoeker)* werd om tijdsredenen in 1983 eveneens niet meer voortgezet.

Het op tabel 6.3 weergegeven registratie- en evaluatiesysteem bleef gehandhaafd. Slechts kleine wijzingen werden aangebracht:

- *Het statistiekformulier* werd zodanig veranderd dat patiënten dit zelf konden invullen.

- *Het diagnoseformulier* werd uitgebreid met een beoordeling van de intake-criteria. Evaluatie hiervan bij ontslag bleek een te vertekenend beeld te geven. Wanneer bij ontslag blijkt, dat een behandeling niet effectief verlopen is,

bestaat het risico dat dit (mede) in de evaluatie van de intake-criteria tot uitdrukking kan komen: Deze patiënt had hier ook eigenlijk niet opgenomen moeten worden. Het diagnoseformulier wordt reeds binnen een maand na opname ingevuld, waardoor van deze vertekening minder sprake zal zijn.

Met het aldus gewijzigd informatiesysteem werd sinds 1983 de programma-evaluatie voortgezet.

6.6 HET EINDE VAN P.O.P.E.

Programma-evaluatie dient een *permanent onderdeel van de organisatie* te zijn: het levert feedback op die voor het beleid van belang c.q. noodzakelijk is. Althans er wordt van uitgegaan dat effectiviteitsgegevens per definitie beleidsrelevant zijn. In dat opzicht zou deze paragraaf dus niet geschreven behoeven te worden. P.O.P.E. had echter naast de doelstelling om deze informatie te verschaffen ook enkele *onderzoekstechnische doelstellingen*. De belangrijkste hiervan werden in de periode 1980-1984 gerealiseerd. Hoewel hierna de informatieverzameling werd afgeslankt, werd deze toch nog te omvangrijk gevonden. Ook andere onderdelen zoals de vertrouwelijkheid van de ontslagvragenlijst wenste men graag te veranderen, zodat o.a. de eigen therapeut deze zou kunnen inzien. Aan welke informatie wél behoefte was, werd niet geëxpliciteerd. Daarnaast bleek dat de feedbackfunctie van de informatie die periodiek werd gerapporteerd, niet (optimaal) werd vervuld. Tevens ontstond de wens om het evaluatieproject geheel in eigen beheer te nemen.

Om organisatorische redenen werd eind 1984 het samenwerkingsverband tussen Overwaal en de Katholieke Universiteit Nijmegen herzien; o.a. de opleiding van assistenten op Overwaal werd gestopt en de hoogleraar psychiatrie was niet meer kwalitaté qua directeur van Overwaal. Vooruitlopend hierop wenste Overwaal de detachering van de researchpsycholoog te beëindigen, en keerde deze terug naar de afdeling psychiatrie.

In 1985 werd het project nog administratief gecontinueerd, terwijl onderzoek startte naar proces-evaluatieve thema's door een researchpsycholoog die uit het eigen budget kon worden bekostigd.

In 1986 nam een psychotherapeut de supervisie van het evaluatieproject op zich. Door hem werd de evaluatie procedure sterk ingekort; onder meer vervielen vijf van de zes follow-up vragenlijsten. In nieuw onderzoek kan nagegaan worden of met deze vereenvoudigde evaluatie de doelstelling van het project: *een objectief informatiesysteem ontwikkelen dat het curatieve functioneren van de instelling permanent evalueert en dusdanige gegevens verschaft, dat deze feedback het functioneren kan verbeteren*, eveneens worden gerealiseerd. De wijze waarop P.O.P.E. is beëindigd, heeft vanzelfsprekend ook betekenis voor de functie die programma-evaluatie vervulde en de positie van de evaluator (zie hfst. 11.3).

Echter, zoals in paragraaf 6.2 is gesteld, reikte de doelstelling van P.O.P.E. verder dan Overwaal, het ontwikkelde evaluatiesysteem zou ook voor andere instellingen bruikbaar moeten zijn, te beginnen met de afdeling Psychiatrie van het Radboudziekenhuis. Dit was immers het uitgangspunt bij de detachering in 1976. Na een proefperiode in een organisatorisch betrekkelijk eenvoudige kliniek als Overwaal zouden de opgedane ervaringen en ontwikkelde instrumenten in het eigen instituut moeten worden gebruikt. Deze doelstelling ligt momenteel binnen bereik: onder leiding van Prof.dr. G.J. Zwanikken bestaat er een uitgesproken behoefte aan periodieke informatie over tal van aspecten van het functioneren van de afdeling, zowel van de polikliniek, de dagkliniek als de kliniek. De ervaringen van P.O.P.E. worden hierbij in extenso aangewend, waarmee het uiteindelijke doel inderdaad gerealiseerd wordt. Wellicht zullen nog meer instellingen in de toekomst volgen. Via een wat lange (om-)weg wordt het uiteindelijke doel toch bereikt.

Hoofdstuk 7:

De kwantitatieve gegevens geregistreerd

7.1 INLEIDING.

Zoals in hoofdstuk 6 beschreven is, verliep Project Overwaal Programma Evaluatie in een viertal chronologisch opeenvolgende fasen. Gestart werd met de registratie van de meer objectieve patiëntengegevens vanaf 1974. Deze registratie bleef gedurende de gehele looptijd van het project bestaan. Zo ontstond een gegevensbestand van ruim 800 opgenomen patiënten. Dit bevatte primair zgn. demografische variabelen, zoals geboortedatum, geslacht, burgerlijke staat, opleiding, beroep, bron van inkomsten, e.d., maar ook gegevens die verband houden met de opname en behandeling: opname- en ontslagdata, voorgeschiedenis voorafgaande hulpverlening, duur van de klachten, diagnose (I.C.D.-9) met de ernst ervan bij opname en ontslag, enz. Bij de uitbreiding van de informatie naar de periode vóór opname en ná ontslag werden ook gegevens verzameld ten aanzien van procedurele zaken zoals: duur wachtperiode (wachttijdstip) voor opname, ontslagprocedure en regeling van de voortgezette hulpverlening. Sinds 1981 werden ook de verstrekte psychofarmaca systematisch vastgelegd, zodat hiervan eveneens overzichten werden verkregen. Deze werden jaarlijks in het jaarverslag vermeld.

In dit hoofdstuk worden deze registratieve gegevens gepresenteerd, waarbij met name gekeken wordt naar mogelijke ontwikkelingen of veranderingen in de loop der jaren of afwijkingen van andere (landelijke) of verwachte gegevens. In par. 7.2 worden de demografische variabelen van de op Overwaal behandelde patiënten van 1974 tot en met 1984 beschreven, voor zover mogelijk in relatie met landelijke gegevens. Verder worden in par. 7.3 de gegevens betreffende de problematiek (diagnose) weergegeven en in par. 7.4 komt de aard en de duur van de voorafgaande hulpverlening aan bod.

Deze registratiegegevens worden hier bivariaat weergegeven: de betrokken parameter gerelateerd aan het opname- of ontslagjaar van de patiënten. Zonodig worden de ontslagjaren gegroepeerd ter vermijding van te grote en onoverzichtelijke tabellen.

7.2 ALGEMENE REGISTRATIE GEGEVENS 1974 T/M 1984.

Over het nut en de betekenis van algemene registratiegegevens lopen de meningen uiteen. Hier kunnen twee gebruiksdoelen onderscheiden worden: ten behoeve van onderzoek of bestemd voor beleid. Deze twee doelen zijn niet strijdig met elkaar. Een goede registratie van (primaire) patiëntenkenmer-

ken is een noodzakelijke conditie voor onderzoek: de doorgaans selecte onderzoekspopulatie dient beschreven te worden teneinde de generalisatiemogelijkheden te kunnen beoordelen; vergelijking met minder selecte, algemenere of eventueel landelijke gegevens kunnen hierbij behulpzaam zijn. Het nut van een dergelijke registratie wordt niet betwist. Anders geldt dit voor de betekenis van registratie voor het beleid. "Registratiegegevens zouden voor de beleidsmakers een kompasfunctie kunnen vervullen als en voorzover zij het mogelijk zouden maken na te gaan of de doelstellingen van de instelling worden gerealiseerd" (Bayer, 1984). Maar zo er al gemakkelijk toegankelijke gegevens worden geregistreerd, dienen zij doorgaans geen ander doel dan opgenomen te worden in het obligate jaarverslag zonder dat dit de door Bayer bedoelde kompasfunctie heeft. Een overzicht van een bepaald gegeven zal wel een signaalfunctie hebben, wanneer dit afwijkt van dat wat men verwachtte of gewenst vindt, maar of dit betekent dat ook het beleid zal worden veranderd teneinde toekomstige resultaten te beïnvloeden is van meer factoren afhankelijk dan het presenteren van dit gegeven alleen. Met name ongewenste consequenties hiervan - vaak financiële - vormen een extrinsieke noodzaak bepaald beleid te herzien. Wanneer deze er niet zijn, zal het signaleren van bepaalde gegevens weinig nut hebben (zie ook hfst. 11).

De hier gepresenteerde overzichten dienen primair ter beschrijving van de onderzoekspopulatie, terwijl opmerkelijke ontwikkelingen c.q. afwijkingen van landelijke gegevens aangestipt worden. Zoals in hfst. 6 is vermeld, werden over 1974 t/m 1976 de zogenaamde statistiekgegevens met terugwerkende kracht uit de status overgenomen. Dit bleek slechts ten dele mogelijk. Vanaf medio 1976 t/m 1982 werden deze gegevens door de onderzoeker genoteerd in een gesprek met de patiënt in de eerste week van zijn opname. Vanaf 1983 vult de patiënt zelf een biografische vragenlijst in. Aldus zijn van 857 patiënten deze statistiekgegevens verzameld. In tabel 7.1 staat een overzicht van het aantal opgenomen en ontslagen patiënten naar ontslagjaar. Het aantal ontslagen patiënten kan afwijken van het aantal opgenomenen doordat patiënten het vorige jaar zijn opgenomen c.q. het volgend jaar worden ontslagen.

Tabel 7.1 Aantal opgenomen en ontslagen patiënten per ontslagjaar

jaar	74	75	76	77	78	79	80	81	82	83	84	tot.
opgen.	55	89	84	74	75	72	69	82	73	82	102	857
ontsl.	55	90	87	75	78	71	67	79	76	81	97	856
herop.	1	7	2	3	2	2	4	8	12	13	16	70

opgen. = opgenomen; ontsl. = ontslagen; herop. = heropnamen; deze zijn in het aantal opnamen begrepen.

Deze tabel laat een wisselend beeld van het aantal opnamen resp. ontslagen zien. In 1974 is er sprake van een startjaar hetgeen het lagere aantal patiënten verklaart. Bij een opnamecapaciteit van 30 bedden en een beoogde verblijfsduur van 4 à 5 maanden (Overwaal was een zgn. A-1 kliniek, zie hfst. 5 en par. 7.2.1) zouden er per jaar 80 à 90 patiënten ontslagen moeten worden. Dit lukte wel in 1975 en 1976, waarna het aantal ontslagen daalde tot 67 in 1980, waarna weer een stijging optrad.

Hoewel het aantal ontslagen patiënten een eenduidig gegeven lijkt, is dat in de praktijk toch minder het geval. Met name wijzigingen in de telprocedure kunnen belangrijke veranderingen veroorzaken, zoals het wel of niet tellen van heropnamen. Tot 1983 werden patiënten die binnen een half jaar - al dan niet intentioneel - weer werden opgenomen, niet opnieuw geteld. Sinds 1983 telt iedere opname, ook na een korte onderbreking bijv. voor een ziekenhuisopname. Over 1983 en 1984 betrof dit 7 resp. 10 patiënten. Uiteraard worden deze patiënten bij hun eerste ontslag niet in de evaluatieprocedure betrokken. Verlaging van het aantal patiënten met deze aantallen blijkt de stijging bijna geheel te nivelleren. Het aantal heropnamen bedraagt sinds 1982 16% van alle opnamen.

Aangezien het ondoenlijk is steeds over 11 jaar afzonderlijk gegevens te presenteren, wordt een verdeling aangebracht in 5 jaargroepen: de eerste drie jaren samen en vervolgens per twee jaar. Alle volgende tabellen zullen in deze jaargroepen worden ingedeeld. Mogelijke trends in de patiëntenpopulatie zijn wellicht via deze groepsindeling beter te herkennen, terwijl incidentele veranderingen eerder vervagen. De groeps grootte bedraagt dan:

periode	aantal opnamen	aantal ontslagen
1974 t/m 1976	228	232
1977 + 1978	149	153
1979 + 1980	141	138
1981 + 1982	155	155
1983 + 1984	184	178
totaal	857	856

7.2.1 Opnameduur.

Een met het aantal ontslagen patiënten direct gerelateerde maat is de (gemiddelde) opnameduur. Deze is voor instellingen voor klinische zorg een cruciaal gegeven en een twistappel tussen de medewerkers en de bestuurlijke c.q. financierende instanties. De ene partij wenst de duur te verlengen om zodoende maximale hulp te kunnen geven, terwijl de andere partij deze om financiële redenen zo kort mogelijk wil houden teneinde zoveel mogelijk patiënten van de hulpverlening te kunnen laten profiteren c.q. het aantal bedden per regio zo klein mogelijk te houden. Voor Overwaal gold dat gestreefd werd naar een gemiddelde opnameduur van 4 à 5 maanden,

waarbij het aantal patiënten dat een opnameperiode van twee kwartaaldata te boven ging, zo klein mogelijk gehouden moest worden. Voor patiënten in wier opnameperiode slechts een kwartaaldatum valt (zgn. A1-patiënten) wordt een hogere verpleegprijs vergoed dan wanneer een patiënt meerdere kwartaaldata opgenomen is (zgn. A2-patiënt).

In tabel 7.2 staat een overzicht van de gemiddelde verblijfsduur per jaargroep. Wordt het streefgetal op 4,5 maanden of 140 dagen gesteld, dan blijkt dat over de totale periode dit gemiddelde gehaald is. Alleen in de 1e en 3e periode is het gemiddelde boven dit niveau. De betrekkelijk constante standaarddeviatie geeft een grote spreiding van de verblijfsduur te zien van circa 3 maanden. Een nadere beschouwing van de verdeling per maand is derhalve zinvol.

Tabel 7.2 Gemiddelde en standaarddeviatie (sd) van de opnameduur in dagen

jaren	74/76	77/78	79/80	81/82	83/84	tot.
gem.	145,2	140,2	152,3	134,3	126,0	139,6
sd.	100,7	92,0	86,5	81,8	87,6	84,4
N	232	153	138	155	178	856

Tabel 7.3 Opnameduur per jaargroep in maanden

duur	74/76		77/78		79/80		81/82		83/84		totaal	
	N	%	N	%	N	%	N	%	N	%	N	%
>1 mnd	32	14	15	10	6	4	14	9	26	15	93	11
1 - 2	22	10	11	7	11	8	14	9	17	10	75	9
2 - 3	26	11	17	11	14	10	19	12	26	15	102	12
3 - 4	26	11	26	17	24	17	27	17	29	16	132	15
4 - 5	29	13	20	13	21	15	21	14	19	11	110	13
5 - 6	19	8	29	19	19	14	21	14	21	12	109	13
6 - 7	21	9	13	9	11	8	15	10	16	9	76	9
7 - 8	16	7	6	4	12	9	9	6	9	5	52	6
8 - 9	14	6	5	3	4	3	5	3	3	2	31	3
9 - 10	8	3	6	4	7	5	3	2	5	3	29	3
10 - 11	4	2	0	0	5	4	4	3	2	1	15	2
11 - 12	7	3	1	1	1	1	1	1	2	1	12	1
>1 jaar	8	3	4	3	3	2	2	1	3	2	20	2
totaal	232		153		138		155		178		856	

Uit tabel 7.3 blijkt, dat 11% en 9% slechts resp. een en twee maanden zijn opgenomen. Gelet op de beoogde behandelingsduur zal hier vaak sprake zijn van een te korte opname c.q. voortijdig ontslag (drop-out). Dit verschijnsel doet zich met name voor in de periode 83/84 waarin 25% van alle patiënten een behandelingsduur van minder dan twee maanden heeft. In par. 9.5 wordt nagegaan in hoeverre hier inderdaad sprake is van een voortijdig ontslag en/of ineffectieve opname.

Naast een absoluut criterium voor wat betreft de opnameduur van 4,5 maanden kan ook een vergelijkend criterium gezocht worden. De beste maatstaf zou dan zijn klinieken voor psychotherapie die patiënten met dezelfde diagnoses opnemen. Een dergelijke vergelijking is niet mogelijk omdat hiervan geen landelijke gegevens voorhanden zijn. Wel beschikt de Geneeskundige Hoofdingspectie voor de Geestelijke Volksgezondheid over een gegevensbestand dat bekend is als Patiëntenregistratie Intramurale Geestelijke Gezondheidszorg (P.I.G.G. zie ook hfst. 1.6). Dit registratiesysteem bevat de algemene psychiatrische ziekenhuizen. De best mogelijke vergelijkende cijfers zijn die, welke betrekking hebben op patiënten die in een dergelijke instelling zijn opgenomen met een diagnose die overeenkomt met de diagnoses die ook aan de patiënten van Overwaal is toegekend. Hiervoor zijn die patiënten uit de P.I.G.G.-overzichten gelicht die gerubriceerd stonden met de diagnose "neurotic conditions". Verschillen in diagnosestelling en behandelingsaard maken dat dit geen optimale vergelijking kan zijn, maar - zoals gezegd - de best mogelijke op dit moment. Tabel 7.4 geeft een vergelijkend overzicht tussen 4 jaargroepen van Overwaal (t/m 1982) en de landelijke gegevens.

Tabel 7.4 Overzicht opnameduur 1974 t/m 1982 (in %) van Overwaal (Ov) en Algemene Psychiatrische Ziekenhuizen (APZ)

opnameduur	74 / 76		77 / 78		79 / 80		81 / 82		totaal	
	Ov	APZ	Ov	APZ	Ov	APZ	Ov	APZ	Ov	APZ
< 1 mnd.	*	*	*	*	4	30	9	29	*	*
1 < 3 mnd.	35	56	28	58	35	27	38	27	34	57
3 < 6 mnd.	32	20	49	20	37	19	38	18	38	19
6 < 12 mnd.	30	14	21	13	22	15	22	14	25	14
> 12 mnd.	3	10	3	9	2	9	2	11	3	10

* niet uitgesplitst, bij de categorie 1 < 3 maanden geteld.

Uit tabel 7.4 is af te leiden dat Overwaal gemiddeld langer behandelt; meer dan de helft van de in APZ opgenomen

patiënten zijn binnen 3 maanden na opname weer ontslagen, terwijl bij Overwaal dit ca. 35% is. Aangezien bij de APZ reeds 30% binnen een maand ontslagen is, en dit voor Overwaal minder dan 10% is, waarbij dan vaak gedacht wordt aan een ineffectieve opname, kan geconcludeerd worden dat de doelstelling van de opname in Overwaal en de APZ verschillend is, in ieder geval ten aanzien van de diagnosegroep "neurotic conditions". Uit ander onderzoek (V.d.Hout, 1984) blijkt overigens dat de korte opnamen in Algemene Psychiatrische Ziekenhuizen eveneens niet effectief zijn. Kennelijk is in deze instellingen het ontslagbeleid op andere gronden gebaseerd dan op Overwaal, waar gestreefd wordt naar een opnameduur van 4 maanden.

7.2.2 Demografische variabelen.

De geregistreerde demografische variabelen bij opname zijn: geslacht, leeftijd, burgerlijke staat, beroep, sociaal niveau, opleiding, bron van inkomsten, woonsituatie. Deze kenmerken worden hier in tabelvorm weergegeven met, waar mogelijk, de vergelijking met de landelijke APZ-gegevens.

- Geslacht.

In tabel 7.5 staat de verdeling naar geslacht vermeld met de APZ-gegevens.

Tabel 7.5 Overzicht van de verdeling naar geslacht van patiënten die van 1974 t/m 1984 zijn opgenomen vergeleken met APZ-gegevens (in perc.).

	74/76		77/78		79/80		81/82		83/84		totaal	
	Ov	APZ	Ov	APZ	Ov	APZ	Ov	APZ	Ov	APZ	Ov	APZ
man	50	36	49	36	38	35	36	35	33	*	39	36
vrouw	50	64	51	64	62	65	64	65	67	*	61	64

* geen gegevens bekend

Deze tabel geeft een opmerkelijk beeld te zien van de verdeling mannen-vrouwen. Voor Overwaal verloopt dit van een gelijke verdeling in de eerste jaargroep naar een verdeling van 1 op 2 in 1983 en 1984. Dit verschil is statistisch significant ($\chi^2 = 9.7$ $p = .02$). Bij de APZ-cijfers is het percentage vrouwen resp. mannen over alle jaren constant: 64% om 36%. Niet bekend is of het aanbod van patiënten voor Overwaal of de selectiecriteria bij intake veranderd zijn. Een nadere analyse van het intakeproces zou hierop antwoord kunnen geven. Of een dergelijke verandering in de opgenomen

populatie ook consequenties heeft voor behandelingseffecten e.d. zal moeten blijken.

- Leeftijd.

De leeftijdsopbouw van de opgenomen patiënten staat weergegeven in tabel 7.6. De gemiddelde leeftijd is begin 30 jaar.

Tabel 7.6 Leeftijden van de opgenomen patiënten.

leeftijd	74/76		77/78		79/80		81/82		83/84		totaal	
	N	%	N	%	N	%	N	%	N	%	N	%
- 24	43	19	25	17	22	16	33	21	38	21	161	19
25 - 34	94	41	67	45	66	47	75	48	82	45	384	44
35 - 44	66	28	34	23	34	24	31	20	41	22	206	24
45 -	29	12	23	15	19	13	16	10	23	12	110	13
gem.	33.4		33.4		33.1		31.9		32.2		32.8	
sd	9.0		9.4		9.1		8.8		9.0		9.1	

Bijna de helft van alle patiënten zijn tussen de 25 en 35 jaar oud. De minimumleeftijd is 17 jaar en de oudste patiënt was 60 jaar. Over de jaren zijn er weinig veranderingen in de leeftijdsverdeling. De gemiddelde leeftijd heeft een lichte tendens tot daling.

- Burgerlijke staat.

De burgerlijke staat van de opgenomen patiënten blijkt een duidelijke verandering te ondergaan ten aanzien van de verdeling gehuwd-ongehuwd. Was dit in 1977/78 63% om 28%, in 1983/84 is deze gelijk: 45% om 45%. Weduwstaat en gescheiden blijft evenwel constant: 2% resp. 8% (zie tabel 7.7). Deze verandering houdt zeker verband met de maatschappelijke wijziging van een toenemende samenwoonsituatie zonder gehuwd te zijn. Dit blijkt ook uit de variabele: woonsituatie (zie tabel 7.8).

De landelijke APZ-gegevens laten een overeenkomstige daling van het percentage gehuwden zien. Het percentage ongehuwden blijft hier evenwel constant, terwijl weduwstaat en gescheiden verdubbeld. Met name is het relatief hoge percentage weduwe/weduwnaar opmerkelijk.

Tabel 7.7 Burgerlijke staat van de opgenomen patiënten van Overwaal (O.) en landelijke APZ (A.) (in %)

burg.staat	74/76 O. A.	77/78 O. A.	79/80 O. A.	81/82 O. A.	83/84 O. A.	totaal O. A.
ongetrouwd	39 31	28 31	33 30	38 31	45 *	37 31
getrouwd	51 56	63 52	54 49	55 46	46 *	53 51
weduwe/-naar	2 8	2 11	3 13	1 15	2 *	2 12
gescheiden	8 4	7 6	10 8	6 8	7 *	8 6

* = nog geen gegevens beschikbaar.

- Woonsituatie.

De woonsituatie bij opname staat in tabel 7.8 vermeld. Hieruit blijkt dat deze weinig verandering ondergaat sinds 1977. De gegevens over 1974-1976 ontbreken. Onder andere blijkt dat de categorie samenwonend/eigen gezin heel constant blijft en ruim 60% bevat, waarmee de opmerking t.a.v. de burgerlijke staat (zie boven) wordt bevestigd.

Tabel 7.8 Woonsituatie van de opgenomen patiënten

	77 / 78		79 / 80		81 / 82		83 / 84		totaal	
	N	%	N	%	N	%	N	%	N	%
alleen	28	19	36	26	40	26	33	18	137	22
oud. gezin	23	15	17	12	22	14	24	13	86	14
eigen gezin	95	64	86	61	90	58	114	62	385	61
elders	3	2	2	1	2	1	13	7	20	3
totaal	149		141		154		184		628	

- Sociaal economisch niveau.

Voor de rubricering van het sociaal economisch niveau wordt gebruik gemaakt van de beroepenklapper van het I.T.S. te Nijmegen (I.T.S., 1975). Een apart probleem doet zich voor bij de patiënten die als beroep "huisvrouw" opgeven. Hierin voorziet de beroepenklapper niet. Indelen volgens het beroep van de partner is niet altijd gewenst of mogelijk. Het indelen bij een van de bestaande categorieën, bijv. "geschoolde arbeid" of "kleine zelfstandige" leek evenmin een juiste keuze en zou tot vertroebeling kunnen leiden. Daarom is "huisvrouw/-man" als een aparte categorie opgenomen. In tabel 7.9 staan de gegevens over de periode 1977-1984 vermeld.

De verdeling over de diverse categorieën blijft over de jaren tamelijk constant: ruim de helft van de patiënten

**Tabel 7.9 Sociaal economisch niveau van de patiënten
opgenomen van 1977-1984.**

s.e.-niveau	77 / 78		79 / 80		81 / 82		83 / 84		totaal	
	N	%	N	%	N	%	N	%	N	%
ongeschoold	4	3	4	3	5	3	13	7	26	4
geschoold	15	10	23	16	18	12	29	16	85	14
lagere emp.	44	30	51	37	54	35	73	40	222	36
kleine zelf.	7	5	1	1	4	3	2	1	14	2
middelbaar	21	14	19	14	13	9	15	8	68	11
hoger	22	15	13	9	17	11	18	10	70	11
huisvrouw	34	23	28	20	41	27	33	18	136	22
totaal	147		139		152		183		621	

behoren tot de drie laagste economische klassen. Bijna een kwart is huisvrouw en de drie hoogste niveaus bevatten eveneens bijna 25%.

- Opleidingsniveau.

Bij opname werd eveneens gevraagd naar de laatst gevolgde (dag)opleiding, ook als deze niet was afgemaakt. Uit tabel 7.10 blijkt dat bepaalde opleidingsniveaus sterk wisselen. Zo blijkt het percentage patiënten met alleen lagere school te wisselen van 13% in groep 2 en 5 naar 4% in groep 4. Een lagere beroepsopleiding komt bij ca. 25% van alle patiënten voor en is vrij stabiel. Het percentage patiënten met MBO stijgt van 17% naar 21%. Van de overige niveaus blijft alleen het percentage patiënten met een universitaire opleiding redelijk stabiel met ca. 6%; alle overige opleidingen wisselen per jaargroep sterk. Met name de periode 1981 en 1982 wijkt sterk af.

Tabel 7.10 Opleidingsniveau opgenomen patiënten.

	77 / 78		79 / 80		81 / 82		83 / 84		totaal	
	N	%	N	%	N	%	N	%	N	%
LO	19	13	8	6	6	4	24	13	57	9
LBO	39	26	35	25	44	28	41	22	159	25
MBO	25	17	27	19	31	20	38	21	121	19
HBO	22	15	24	17	31	20	20	11	97	16
MVO	15	10	19	14	9	6	28	15	71	11
HVO/VWO	18	12	17	12	27	17	18	10	80	13
Un. + TH	10	7	9	6	7	5	15	8	41	7
totaal	148		139		155		184		626	

Bijna 50% van alle patiënten heeft een LO-, LBO- of MVO-opleiding.

- Bron van inkomsten.

De bron van inkomsten voor opname is gesplitst in vijf groepen: werk, werkeloos (WW, RWW, Bijstand), ziek (ziektewet of WAO), pensioen/studerend e.d. en geen eigen inkomen (incl. alimentatie). In tabel 7.11 staan de gegevens over de patiënten van 1977 t/m 1984. Ook deze tabel laat soms grote schommelingen zien. Zo wisselt het percentage patiënten zonder eigen inkomen tussen 18% in periode 3 en 34% in periode 4.

Tabel 7.11 Bron van inkomsten van de opgenomen patiënt.

jaren	77 / 78		79 / 80		81 / 82		83 / 84		totaal	
inkomsten	N	%	N	%	N	%	N	%	N	%
werkzaam	43	29	36	26	23	15	49	27	151	24
werkeloos	9	6	8	6	12	8	27	15	56	9
ziek	52	35	58	41	64	42	60	33	234	37
stud./pens	7	5	13	9	7	5	14	8	41	7
geen eigen	36	24	25	18	48	31	34	19	143	23
totaal	147		140		154		184		625	

Vanzelfsprekend is de groep met een ziekte- of WAO-uitkering het sterkst vertegenwoordigd met ca. 37% van alle patiënten.

- Arbeidssector. Een opmerkelijk stabiel gegeven is de arbeidssector, waarin de opgenomen patiënten (voor het laatst) gewerkt hebben. De sector "verzorgende beroepen" is met 27% het sterkst vertegenwoordigd, hetgeen veroorzaakt wordt door de huisvrouwen die hierin eveneens zijn ingedeeld. Uit de administratieve sector komt 21% van de patiënten, 19% uit handel en industrie, 15% uit het onderwijs (incl. de studerenden), 13% uit medische of sociale beroepen en tenslotte 5% uit de agrarische, bestuurlijke of wetenschappelijke sector.

- Samenvatting.

Hiermee zijn de demografische variabelen beschreven. Samengevat kan gesteld worden, dat de Overwaalpopulatie kan worden gekenschetst als: tussen de 25 en 35 jaar oud, vaker een vrouw dan een man (in de verhouding 2 : 1), gehuwd en/of samenwoont (60%), uit de drie laagstesociale klassen afkomstig (55%) of een huisvrouw (22%); de bron van inkomsten is vaak een ziekte- of W.A.O.-uitkering of er is geen eigen inkomen (huisvrouw).

Nog niet eerder werd vermeld dat 70% van alle patiënten in het R.K.-geloof is opgevoed, 20% in een ander geloof en 10% niet in een geloof is opgevoed. Van hen die wel in een geloof zijn opgevoed, is nog ca. 40% praktizerend of beleidend godsdienstig.

7.3 PROBLEMATIEK VOOR OPNAME.

In 7.2 werden de zgn. demografische variabelen van de opgenomen patiënten beschreven. In deze paragraaf komen de gegevens aan de orde die verband houden met de problematiek bij opname. Dit betreft de diagnose bij opname, de duur van de klachten, de totale duur van de hulpverlening voor opname, e.d. Deze gegevens zijn minder "hard" dan de beschreven variabelen in 7.2. De diagnose is een subjectieve abstractie van de problematiek zoals deze door de psychotherapeut geïnterpreteerd wordt. De gegevens over duur van de klachten en de voorafgaande hulpverlening zijn bij opname genoteerd en kunnen vertekend zijn. Sommige patiënten hebben een dermate omvangrijk arsenaal van voorafgaande hulpverlening gehad, dat dit niet meer accuraat reproduceerbaar is. Een gegeven als de duur van de klachten is dikwijls eveneens niet exact traceerbaar: klachten ontstaan doorgaans heel geleidelijk zonder dat de betrokkene zich hiervan bewust is. Het aangeven van een begintijdstip is dan niet precies mogelijk, maar zal slechts een benadering zijn. Van 1977 tot en met 1982 werden deze gegevens door de onderzoeker in een gesprek met de pas opgenomen patiënt genoteerd. Zodoende kon enige waarborg geschapen worden voor een juiste interpretatie en beantwoording van deze vraagstellingen. Als bijv. een patiënt antwoordde zijn hele leven reeds last van zijn klachten te hebben, werd niet zijn leeftijd als duur van de klachten genoteerd, maar werd doorgevraagd naar het tijdstip dat betrokkene zich bewust was van zijn klachten. Wanneer patiënten dit zelf invullen, zal beantwoording minder uniform geschieden.

- Diagnose.

Sinds het invoeren van de P-registratie volgens P.I.G.G. wordt door de psychotherapeut binnen 1 maand na opname een diagnose volgens I.C.D.-9 gegeven (WHO, 1978). In totaal mogen er 5 diagnoses gegeven worden, welke in volgorde van belangrijkheid worden genoteerd. In tabel 7.12 worden de diagnoses van de 4 laatste jaargroepen gepresenteerd.

Hieruit blijkt, dat steeds ca. 60% van alle gestelde diagnoses gerangschikt kan worden onder de neurosen. Hierbij komen "fobie" (22%), "dwang" (12%) en "overige" (11%) het meeste voor. Bij de overige 40% wordt het meest een sociale diagnose, de zgn. V-codes gegeven (19%). Er zijn weinig verschillen tussen de jaargroepen in de aard van de totale diagnoses. Bij nadere beschouwing blijkt er wel een verandering

Diagnose:	77 /		78 /		79 /		80 /		81 /		82 /		83 /		84 /		totaal		
	1e	T	%		1e	T	%		1e	T	%		1e	T	%		1e	T	%
Neurose:																			
- angst	16	19	5		14	16	3		16	25	6		10	17	4		56	77	4
- hysterie	4	7	2		7	9	2		5	6	1		4	4	1		20	26	2
- fobie	43	96	25		47	112	24		36	68	16		44	107	22		170	383	22
- dwang	18	34	9		21	50	11		27	49	12		36	73	15		102	206	12
- depressie	18	24	6		14	32	7		11	23	5		10	20	4		53	99	6
- neurastenie	4	12	3		2	3	1		3	5	1		2	5	1		11	25	1
- depersonal.	-	1	0		1	1	0		1	3	1		-	-	0		2	5	0
- hypochondrie	1	2	1		1	3	1		1	1	0		-	-	0		3	6	0
- andere	2	3	1		-	-	0		2	2	1		-	-	0		4	5	0
- overige	12	26	7		17	51	11		22	53	13		26	67	14		77	197	11
Totaal neurose	118	224	60		124	277	59		24	235	56		132	293	60		498	1029	58
Persoonlijkheid	12	22	6		9	21	4		6	11	3		12	30	6		39	84	5
Sexueel	1	7	2		-	21	4		1	14	3		1	10	2		3	52	3
Verslaving	2	7	2		2	13	3		1	10	2		-	9	2		2	39	2
Funktioneel	4	14	4		2	16	3		-	9	2		-	6	1		6	45	3
Eetstoornis	1	4	1		6	13	3		5	12	3		7	14	3		19	43	2
Overig spec.	1	4	1		-	12	2		2	17	4		1	11	2		4	44	3
Aanpassing	5	19	5		5	15	3		7	16	4		8	17	4		25	67	4
Depressie	1	2	1		2	7	2		1	3	1		1	4	1		5	16	1
Sociaal	3	72	19		2	77	16		2	91	22		2	95	19		9	335	19
Psychotisch	1	1	0		-	-	0		1	1	0		3	3	1		5	5	0
Totaal overige	31	152	40		28	195	41		26	184	44		35	199	40		120	730	42
Totaal totaal	376				472				419				492				1759		

Tabel 7.12 Diagnosen van de patiënten opgenomen
van 1977 t/m 1984

te zijn bij de eerste diagnose. Het aantal patiënten met dwang als eerste diagnose stijgt van 12% in 77/78 naar 22% in 83/84. Een dergelijke grote verschuiving doet zich bij geen andere eerste diagnose voor.

- Duur van de klachten.

Bij opname wordt aan de patiënt gevraagd hoelang deze zich reeds bewust is van zijn problemen; gevraagd werd naar het tijdstip dat hij zich realiseerde dat er problemen waren, dat hij er met anderen over sprak of dat anderen hem hierop attenderden. Het gegeven antwoord werd gecodeerd in periodes. In tabel 7.13 staan de resultaten in jaargroepen vermeld.

Tabel 7.13 Hoelang is de patiënt zich reeds bewust van zijn problemen bij opname.

	77/78		79/80		81/82		83/84		totaal	
	N	%	N	%	N	%	N	%	N	%
niet	5	3	-	0	-	0	1	0	6	1
- ½	6	4	4	3	5	3	16	9	31	5
½ - 1 jr	10	7	5	4	6	4	10	5	31	5
1 - 3 jr	19	13	21	15	23	15	26	14	89	14
3 - 6 jr	49	33	36	26	32	21	47	26	164	26
> 6 jr	60	40	74	53	86	57	83	45	303	49
totaal	149		140		152		183		624	

Zoals uit tabel 7.13 blijkt, is bijna de helft van alle patiënten zich reeds meer dan 6 jaar bewust van zijn problemen. Dit percentage steeg van 40 naar 57% in de periode 1977 tot 1982. Over de laatste twee jaar daalde dit weer naar 45% en steeg met name de groep patiënten die aangeeft zich slechts kort bewust te zijn van hun problemen.

- Mate en duur van dysfunctioneren.

Bij opname werd eveneens gevraagd sinds wanneer de patiënt niet meer in staat was door de problemen waarvoor hij nu werd opgenomen zijn werk of bezigheden te doen en in welke mate hij hierin belemmerd werd. In tabel 7.14 zijn de resultaten weergegeven.

Uit tabel 7.14 blijkt, dat ca. 40% niet of slechts kortdurend dysfunctioneert door hun problemen, terwijl bijna 30% reeds meer dan 3 jaar erdoor belemmerd wordt. Dit percentage is sinds 1977 gestegen van 20 naar 32%.

Tabel 7.14 Duur van dysfunctioneren bij opname.

	77 / 78		79 / 80		81 / 82		83 / 84		totaal	
	N	%	N	%	N	%	N	%	N	%
niet	36	25	23	16	20	13	33	18	112	18
- ½	40	27	27	19	39	26	37	20	146	23
- 1 jr	14	10	16	11	14	9	16	9	61	10
- 3 jr	27	18	29	21	37	24	38	21	132	21
- 6 jr	15	10	24	17	22	14	28	15	89	14
> 6 jr	14	10	21	15	21	14	32	17	88	14
totaal	146		140		153		184		623	

Gevraagd naar de ernst van het belemmerd worden bij de dagelijkse bezigheden op het moment van opname antwoordt 2/3 van alle patiënten "tamelijk veel" tot "zeer veel - ik kan eigenlijk niets doen". Dit is over alle jaargroepen hoege-naamd gelijk. Er doet zich wel een verschuiving voor in de hoogste gradatie van dysfunctioneren: het percentage patiënten dat zegt eigenlijk niets meer te kunnen doen, daalt van 26% in de 2e jaargroep naar 14% in de 5e jaargroep. De gegevens in de een na hoogste gradatie zijn hiermee precies tegengesteld, van 15% in de 2e jaargroep naar 33% in de 5e jaargroep, zodat gemiddeld de mate van dysfunctioneren daalt.

- Duur voorgaande hulpverlening.

Een andere maat voor de duur van de problemen is de periode dat een patiënt reeds eerder hulp heeft gehad. Over het algemeen is deze periode korter dan de tijd dat hij zich bewust is van zijn problemen.

Tabel 7.15 Hoelang heeft de patiënt reeds hulp gehad.

	77 / 78		79 / 80		81 / 82		83 / 84		totaal	
	N	%	N	%	N	%	N	%	N	%
geen	9	6	1	1	-	-	-	-	10	2
- ½ jr	16	11	8	6	17	12	22	12	63	10
- 1 jr	15	10	11	8	8	5	23	13	57	9
- 3 jr	31	21	38	27	33	22	42	23	144	23
- 6 jr	39	26	38	27	41	28	44	24	162	26
> 6 jr	38	26	43	31	48	33	50	28	179	29
totaal	148		139		147		181		615	

In tabel 7.15 staat vermeld dat ruim 50% reeds meer dan 3 jaar geleden hulp heeft gezocht. Uit een andere vraag blijkt, dat ca. 30% van alle patiënten niet onafgebroken deze hulp hebben gehad. Uit tabel 7.15 is eveneens af te leiden dat er een lichte tendens is tot een minder lange duur van de voorafgaande hulpverlening. Overeenkomstig de periode dat de patiënt zich bewust was van zijn problemen, vertoont ook de duur van de hulpverlening aanvankelijk een lichte stijging, maar in de laatste periode weer een daling.

In par. 7.4 wordt nader op dit onderdeel ingegaan, waarbij vooral gekeken wordt naar de aard - klinisch of ambulante - van deze voorafgaande hulpverlening.

- Verwijzer naar Overwaal.

De verwijzer naar Overwaal is de laatste schakel in het proces van voorafgaande hulpverlening. Redenerend dat een patiënt via het eerste en tweede echelon in een klinische setting wordt opgenomen, zou de verwijzende instantie tot het tweede of het derde echelon moeten behoren. Evenwel speelt in veel gevallen de huisarts een tussenschakel tussen hulpverlenende instanties. Daarnaast komt het ook voor, dat patiënten zich zelf aanmelden. In dat geval wordt doorgaans eveneens de huisarts als verwijzer aangemerkt. Dit is althans het beleid op Overwaal sinds 1980. In tabel 7.16 staat vermeld via welke verwijzers de patiënten op Overwaal werden opgenomen.

Tabel 7.16 De verwijzers en verwijsinstanties

	77/78		79/80		81/82		83/84		totaal	
	N	%	N	%	N	%	N	%	N	%
huisarts/zelf	35	24	46	33	41	27	38	21	160	26
Riagg, e.d.	31	21	46	33	48	31	63	34	188	30
prive praktijk	18	12	5	4	19	12	14	8	56	9
poliklinieken	45	30	31	22	21	14	30	16	127	20
klinieken	6	4	9	6	13	8	23	13	51	8
Overwaal	14	9	4	3	13	8	15	8	46	7
totaal	149		141		155		183		628	

Deze tabel laat zien, dat het RIAGG de grootste verwijzer is met 30%, dan volgt de huisarts met 26%. De poliklinieken van psychiatrische ziekenhuizen of P.A.A.Z. verwijzen 20% van alle patiënten. Hun aandeel daalde van 30% naar 16%. Hiertegenover staat een stijging van de klinische verwijzingen van 4 naar 13%. Hun aandeel is evenwel gering met 8% van het totaal en ongeveer gelijk aan de verwijzingen via privé praktijken van psychiater of psycholoog en de verwijzingen via Overwaal. Deze laatste categorie bevat doorgaans part-

ners van reeds opgenomen patiënten, heropnamen van patiënten die weer met hun psychotherapeut contact opnamen of patiënten die een psychotherapeut van Overwaal kenden.

- Samenvatting.

Ca. 60% van alle patiënten hebben de diagnose "neurose", waarbij fobie en dwang het meeste voorkomen. Bij 44% van de patiënten worden deze als eerste diagnose opgegeven. De problemen bestaan bij bijna de helft van de patiënten meer dan zes jaar. De hulpverlening bestaat bij meer dan 50% reeds langer dan drie jaar. De aard of de ernst van de problemen is niet steeds zo dat dit tot dysfunctioneren leidt: 40% dysfunctioneert niet of slechts kortdurend, terwijl 30% meer dan drie jaar belemmerd wordt. De belangrijkste verwijzers zijn het RIAGG, de huisarts en de psychiatrische poliklinieken. De veranderingen over de periode van 1977 tot 1985 zijn slechts gering: het aantal patiënten met als eerste diagnose "dwang" steeg, de duur van de klachten steeg aanvankelijk, maar daalde in de periode 1983/84 weer, evenals de duur van de hulpverlening vóór opname. Als de in deze paragraaf beschreven variabelen iets zeggen over de ernst van de problematiek bij opname, dan blijkt uit de gepresenteerde resultaten niet dat deze een monotoon stijgende lijn vertoont. Een geleidelijke stijging doet zich op tal van variabelen voor over de periode 1977 tot 1983, waarna er weer een daling optreedt. In hoofdstuk 9 wordt nagegaan welke relaties bestaan tussen deze gegevens en diverse effectmaten.

7.4 HULPVERLENING VOOR OPNAME.

In de vorige paragraaf werd de totale duur tussen het eerste contact met een hulpverlener en de opname op Overwaal vermeld. Deze varieerde van geen eerdere hulpverlening (2%) tot 6 jaar of meer (29%). Deze periode was bij opname reeds in categorieën verdeeld. Echter, deze totale periode is wellicht geen goede indicatie voor de duur van de hulpverlening, aangezien er onderbrekingen van langere of kortere duur kunnen zijn geweest. Daarnaast zegt het niets over de aard van de hulpverlening: deze kan korte gesprekjes met de huisarts zijn geweest tot langdurige en/of meermalige opnamen in psychiatrische ziekenhuizen. Bij opname op Overwaal werd gevraagd naar deze behandelingsvoorgeschiedenis: per hulpverleningsinstantie werd genoteerd wanneer, hoelang en hoe intensief deze behandeling geweest was. Voor onderstaande overzichten werd de aard van de hulpverlening in drie categorieën gesplitst: klinisch, ambulant en overige. Onder "overige" werd verstaan: hulp van de huisarts, de paramedische, alternatieve en paranormale hulpverlening. Deze categorie blijft hier verder buiten beschouwing. Per hulpverleningscategorie werd de totale duur in maanden gesommeerd alsmede het aantal verschillende hulpverleners genoteerd. Niet iedere

patiënt wist zich te herinneren hoelang bepaalde hulp had geduurd, zodat dit gegeven herhaaldelijk ontbreekt. Sinds 1983 vulden patiënten zelf het zogenaamde statistiekformulier in, terwijl dit voordien door de onderzoeker in een gesprek met de patiënt gebeurde. Dit leidde er met name toe dat 25% van de patiënten deze, toch vaak wat moeilijk te memoreren informatie, niet invult. Voorheen werden deze gegevens bij slechts 7% gemist.

Om na te kunnen gaan of met name de intensiteit van de voorafgaande hulpverlening is veranderd in de loop van de jaren, hetgeen een indicatie zou kunnen geven voor de ernst van de problematiek, zijn de volgende onderverdelingen gemaakt:

- geen voorgaande hulp
- een, twee en drie of meer ambulante resp. klinische behandelingen

Naast het aantal verschillende hulpverleners zal ook nog de duur van de hulpverlening (klinisch en/of ambulant) bekeken worden.

Tabel 7.17 Overzicht van de aard en het aantal voorafgaande hulpverlening/-ers.

	77 / 78		79 / 80		81 / 82		83 / 84		totaal	
	N	%	N	%	N	%	N	%	N	%
klinisch										
0	91	64	80	59	86	62	80	58	337	61
1	28	20	31	23	33	24	35	25	127	23
2	13	9	15	11	7	5	17	12	52	9
> 3	10	7	9	7	12	9	6	4	37	7
totaal	142		135		138		138		553	
ambulant										
0	20	14	8	6	12	9	9	6	49	9
1	44	32	50	38	46	33	57	42	197	36
2	40	29	47	36	38	28	34	25	159	29
> 3	35	25	27	20	41	30	37	27	140	26
totaal	139		132		137		137		545	

In tabel 7.17 staat een overzicht van het aantal verschillende hulpverleningen; ca. 40% is al eens eerder opgenomen geweest. Meer dan 50% heeft twee of meer ambulante hulpverleners bezocht. Uiteraard hebben bijna alle patiënten met een klinische hulpverlening ook een of meer ambulante hulpverleners gehad. Er zijn weinig verschillen tussen de jaargroepen. Het grote aantal ontbrekende gegevens over 83/84 kan dit evenwel vertekenen. Het lijkt goed mogelijk, dat patiënten

met een uitgebreid hulpverleningsverleden eerder de gegevens niet verstrekken.

Het voert te ver hier alle combinaties van vormen van hulpverlening weer te geven. De duur van de verschillende vormen van hulp laat zich als volgt beschrijven:

- Patiënten die eenmaal eerder zijn opgenomen geweest, verbleven daar gemiddeld 3 maanden. Waren zij tweemaal opgenomen, dan bedroeg de totale opnameduur gemiddeld 7 maanden. Bij drie of meer eerdere opnamen was deze totale duur 16,9 maanden. Deze categorieën tezamen nemend: als een patiënt eerder was opgenomen, verbleef hij hier gemiddeld 5,9 maanden; 13 patiënten waren reeds meer dan een jaar klinisch behandeld.
- De overeenkomstige cijfers bij ambulante hulp zijn: gemiddeld 11,8, 23,2 en 29,6 maanden bij resp. 1, 2 en 3 of meer verschillende hulpverleners. In totaal hebben 96 patiënten 1 tot 5 jaar ambulante hulp gehad en 13 patiënten zelfs meer dan 5 jaar.
- De totale periode vanaf het eerste hulpverleningscontact tot het moment van opname in Overwaal is gemiddeld 56 maanden. Van de patiënten van wie alle gegevens over de duur van de behandelingen compleet zijn ($N = 200$), is de verdeling van de hulperiodes als volgt: totale periode 42 maanden, klinische periode 2,3, ambulante periode 16,5 en overige hulp gemiddeld 3,4 maanden. Dit betekent, dat in totaal gemiddeld 22,2 maanden hulp ontvangen is en dat 20 maanden geen hulpverlening gekregen is, ofwel bijna de helft van de totale periode.

Samengevat is de diversiteit van voorafgaande hulpverlening zeer groot: van geen enkele hulp tot zeer langdurige klinische en/of ambulante hulpverlening. Er zijn geen duidelijke veranderingen zichtbaar tussen de jaarperiodes. Hoewel dikwijls de indruk bestaat, dat klinische hulp steeds langer wordt uitgesteld c.q. tegengegaan bij grotere professionalisering van de ambulante hulpverlening, is dit bij deze patiëntenpopulatie nauwelijks waarneembaar. Van alle patiënten die ambulante hulp gehad hebben, is de daadwerkelijke duur per jaargroep resp. 16, 18, 21 en 19 maanden.

Hiermee eindigt deze beschrijving van de gegevens van ruim 800 patiënten die over een periode van 10 jaar zijn opgenomen. Uiteraard is een selectie gemaakt uit het uitgebreide gegevensbestand. Daarnaast leent dit bestand zich nog voor nadere analyses. Het voert te ver dergelijke analyses hier te presenteren. In toekomstig onderzoek zal in dit bestand verder naar relaties tussen de variabelen onderling worden gezocht.

Hoofdstuk 8:

De evaluatie van de resultaten

*Succes kent vele vaders, maar
mislukking is een eenzame wees.*
J.F. Kennedy.

8.1 INLEIDING.

Zoals in hoofdstuk 6 vermeld is, heeft het programma-evaluatie project zich in een viertal opeenvolgende fasen voltrokken: een retrospectief follow-up onderzoek (1974-1976), een permanente ontslag en follow-up evaluatie (nov. 1976-1980), een vergelijkende studie tussen meetinstrumenten (1980-1983) en een vereenvoudigde evaluatiefase (1983-1985). De gebruikte meetinstrumenten verschillen in deze vier fasen in meerdere of mindere mate. Het meest uniform zijn de in dit project ontwikkelde vragenlijsten die door de (ex-)patiënt en de behandelaar zijn ingevuld. Getracht is de resultaten over de periode 1976 tot 1985 aan-eensluitend te presenteren, omdat de veranderingen van de vragenlijsten in 1980 hoofdzakelijk uitbreidingen inhielden.

Echter, bij analyse bleek dat vragen die slechts weinig veranderd waren, soms toch grote verschuivingen in de antwoorden te zien gaven (zie par. 8.5). Aangezien dit een onjuist beeld zou geven van de ontwikkeling van het behandelingsresultaat in de tijd, was het noodzakelijk de evaluatiegegevens per fase te presenteren. In de opeenvolgende paragrafen worden de belangrijkste resultaten van de door de onderzoeker geconstrueerde ontslag- en follow-up evaluatie vragenlijsten weergegeven in drie perioden:

- 1974-1976 (alleen retrospectief follow-up onderzoek).
- 1977-1980 (ontslag- en drie follow-up-vragenlijsten).
- 1980-1985 (opname-, ontslag- en follow-up-lijsten).

Voor de twee eerste perioden wordt eveneens gebruik gemaakt van eerdere rapportages (Sijben, 1977c, 1979b). De derde periode is alleen herhaaldelijk intern gerapporteerd. (Sijben, 1984c)

De presentatie van de gegevens is in hoofdzaak descriptief. De bedoeling is een idee te geven welke informatie een dergelijk project op microniveau kan opleveren, en welke rapportage de opdrachtgevers c.q. de medewerkers zouden kunnen verwachten. Een verslag zoals hier gegeven, is periodiek intern gepresenteerd en met de psychotherapeuten en chef de clinique besproken, zoals dat in een programma evaluatie project betaamt: feedback is het eerste doel en dient voor de betrokkenen inzichtelijk en begrijpelijk te zijn. Informatie op vraagniveau voldoet hier aan. De meer gecompliceerde gegevens staan vermeld in hoofdstuk 9.

Bij de start van het evaluatieproject in 1976 werd begonnen met het opzetten van een registratiesysteem (zie hfst. 7) en een retrospectief follow-up onderzoek bij de patiënten die van 1974 tot november 1976 waren ontslagen. Aan deze patiënten werd een vragenlijst toegestuurd, wanneer ze minimaal 3 maanden waren ontslagen. Dit betrof in totaal 213 patiënten. Aan 12 patiënten werd geen vragenlijst verzonden op verzoek van de therapeut (om uiteenlopende redenen), terwijl 13 patiënten twee maal waren opgenomen en uiteraard slechts eenmaal een lijst ontvingen. Aan 188 patiënten werd bijgevolg een vragenlijst verzonden, waarvan er 9 als onbetelbaar terug kwamen. Van de resterende 179 lijsten werden er 125 ingevuld geretourneerd (70%). Over de drie ontslagjaren zijn de responsepercentages: 54, 72 en 77%. Deze responsratio verschilt per opnameduur: patiënten die slechts kort zijn opgenomen (< 1 maand) sturen minder vaak een lijst terug evenals patiënten die langer dan 9 maanden opgenomen zijn geweest. Mannen sturen vaker een lijst terug dan vrouwen. De aantallen per subgroep zijn echter te klein om hier conclusies aan te verbinden.

De belangrijkste uitkomsten van dit follow-up onderzoek worden hier in vraag en antwoord weergegeven:

- Wat gebeurde er met de problemen waarvoor u werd opgenomen?
 - 15%: niet behandeld
 - 40%: geheel/tamelijk goed verdwenen
 - 45%: een beetje/niet verdwenen of slechter geworden.
- Waren er nog andere problemen?
 - 17%: ja, maar ze werden niet behandeld
 - 8%: ja, ze zijn geheel verdwenen
 - 57%: ja, ze zijn een beetje verdwenen
 - 18%: ze zijn niet verdwenen of slechter geworden.

Deze twee eenvoudige effectvragen laten zien, dat het beoordeelde effect: de mate van verdwenen zijn van de problemen, door de patiënten niet erg hoog wordt ingeschat. Hierbij moet worden opgemerkt, dat deze beoordeling geschiedt 3 tot 36 maanden na ontslag. Want als reden voor ontslag geeft 60% op, dat de problemen toen zover over waren dat men het alleen (10%) of met nazorg (50%) weer aan kon.

De vragen naar de status quo op moment van invullen leveren de volgende antwoorden op:

- Hoe voelt u zich nu met betrekking tot de problemen?
 - 17%: problemen zijn over
 - 57%: problemen zijn minder
 - 26%: problemen zijn er nog/ernstiger.

- Verschil tussen opname op Overwaal en nu?
 51%: veel beter
 33%: wel wat beter
 16%: hetzelfde/slechter.
- Bent u nog onder behandeling?
 13%: nee, problemen zijn over
 60%: ja
 27%: nee, maar problemen zijn er nog.

Het interpreteren van deze gegevens is niet eenvoudig, met name door de vraag waar het criterium voor een effectieve behandeling gelegd moet worden. Er is een tamelijk grote mid-dengroep die aangeeft dat de problemen "weinig" of "wel wat" verbeterd zijn. Wanneer deze bij de niet-effectief behandelde patiënten worden gerekend, neemt deze groep bijna 50% van alle patiënten in beslag, een niet erg positief resultaat. Worden deze daarentegen bij de effectieve groep geteld, dan is het effectpercentage ca. 80%. Voor beide berekeningswijzen zijn argumenten aan te voeren. De doelstelling van de kliniek is in deze niet duidelijk.

Naast de effectvragen werden ook vragen gesteld naar de satisfactie van de opname, zowel naar de behandeling en behandelaar, als naar het verblijf en aspecten daarvan zoals huisvesting, e.d. De beoordelingen van deze vragen zijn zeer positief zoals uit onderstaande opsomming blijkt:

- oordeel over psychotherapeut: Ik zag deze tamelijk/heel veel (74%), heeft me tamelijk/heel goed geholpen (75%), het contact was tamelijk/erg goed (81%), het contact was erg/tamelijk prettig (89%);
- de kwalificatie "goed" voor de diverse behandelingsvormen varieert van 35% voor een bepaalde groepstherapie tot 75% voor de individuele psychotherapie, de belangrijkste behandelingsmethode;
- de beoordeling van de verblijfsaspecten is gunstig: de contacten met de medepatiënten worden als erg/tamelijk prettig genoemd (95%), de huisvesting als tamelijk/erg gezellig (61%), de leefregels als tamelijk/erg vrij (97%). Met dit laatste is overigens niet iedere patiënt het eens.

Een derde evaluatie-aspect zijn de procedurele zaken, zoals regeling van ontslag en nazorg:

- tijdstip van ontslag: 54% zegt op tijd ontslagen te zijn, 32% te vroeg en 14% te laat;
- voor 21% was geen nazorg geregeld, één derde van deze patiënten had wel nazorg gewenst.

Als laatste was er een restcategorie van vragen over de situatie op moment van invulling, zoals bron van inkomsten, hulp situatie en -behoefte, ervaren stigmatisering. De resultaten van deze vragen zijn:

- bron van inkomsten (patiënt of partner): 57% werkte, 1% had WW, 32% WAO of ziektewet en 10% was studierend of had pensioen;
- naarmate men langer ontslagen is, daalt het percentage patiënten dat nog behandeling heeft van 76% bij 3-6 maanden na ontslag naar 38% bij 30-36 maanden;

- op het moment van invulling heeft 40% geen hulpverlening, één vierde van hen zou wel graag (weer) hulp hebben. Overigens, van de patiënten die wel hulp hebben, zegt 30% andere hulp te wensen.
- uit de stigmatiseringsvragen blijkt, dat 45% van de patiënten een begripvolle reactie ontvingen t.a.v. de opname, 27% kreeg weinig begrip en 13% geen begrip of zelfs afwijzing. De resterende 15% zegt dat de omgeving niet op de hoogte was en/of niet reageerde (een dergelijke opname wordt soms "geheim" gehouden). Overigens blijkt uit een andere vraag, dat 15% ervaren heeft wel eens te zijn afgewezen bij officiële instanties (12%) of vrienden (3%). Deze resultaten komen redelijk overeen met het onderzoek van Van den Hout. Hij meldt een ervaren etikettering van 26% na 6/8 maanden (Van den Hout, 1985, p. 133). Verschillen in vraagvormen kunnen voor de afwijking van de resultaten verantwoordelijk zijn.

Wordt tenslotte naar verbanden gezocht tussen de antwoorden op de vragen enerzijds en demografische gegevens zoals die in de status vermeld zijn anderzijds, dan blijken deze niet aanwezig. Effect en satisfactie zijn niet gerelateerd aan patiëntkenmerken, noch aan diagnose. Het enige aangetoonde verband is dat tussen effect en behandelaar. Op Overwaal waren steeds 2 arts-assistenten in opleiding gedurende ca. 1 jaar als therapeut werkzaam, onder supervisie van een ervaren psychotherapeut. Wanneer deze arts-assistenten werden vergeleken met de vaste medewerkers, dan bleken tussen deze twee groepen behandelars grote verschillen te bestaan. Op alle eerdergenoemde vragen antwoordden de patiënten die door assistenten werden behandeld, significant negatiever dan bij de vaste medewerkers. Dit gold voor de effectvragen zowel als voor de satisfactie vragen. Zo meldde 28% van de patiënten dat de problemen niet behandeld waren en 25% dat ze hetzelfde waren gebleven of slechter geworden. Bij de vaste staf waren deze uitkomsten resp. 9% en 14%. Ook het niet regelen van nazorg kwam bij assistenten veel vaker voor.

Bovenstaande resultaten zijn binnen het team besproken. Er werd geen aanleiding gezien tot verandering van het behandlingsbeleid. De algemene effectiviteit werd als voldoende geoordeeld, doordat iedere mate van verbetering als gunstig werd ervaren en men derhalve op een effectiviteit van ca. 75% kon terugzien. In de uitsluitend intern gepubliceerde verschillen tussen de twee groepen behandelars werd evenmin aanleiding gezien tot veranderingen. Een dergelijk verschil was immers geheel begrijpelijk. Daarnaast was binnen de groep assistenten een grote heterogeniteit qua beoordeeld effect en hebben ze minder tijd vanwege hun opleiding, e.d. Geconcludeerd werd, dat gezocht zal worden naar de werkelijke oorzaken en getracht zal worden deze op te lossen, "... zodat daardoor Overwaal in zijn geheel en de assistenten in het bijzonder beter beoordeeld zullen worden en dus wellicht beter zullen functioneren" (Sijben, 1977c). In de navolgende 8 jaar moet blijken of dit gerealiseerd kan worden.

Dit retrospectieve follow-up onderzoek heeft aan zijn doelstellingen voldaan: in korte tijd werd ervaring opgedaan met dit type onderzoek, een vragenlijst ontwikkeld, resultaten verwerkt en gerapporteerd aan de staf van Overwaal. Binnen een jaar was dit onderzoek afgerond. De presentatie van de resultaten had vooral tot doel de therapeuten vertrouwd te maken met deze vorm van evaluatie en feedback. De rapportage was heel therapeut-vriendelijk: harde conclusies waren immers niet te trekken, omdat de onderzochte groep patiënten te klein en te heterogeen was, en zich uitstreckte over 3 jaar. Daarnaast was geen criterium bekend of gesteld voor een effectieve behandeling. Op dit moment werd iedere mate van verbetering als effectief gezien. Over de hoogte van het effectiviteitscriterium zijn evenwel nog enige kanttekeningen te maken.

Gelet op het feit, dat deze patiënten zeer intensief (klinisch) behandeld zijn, zou een beter resultaat dan "een beetje" verwacht kunnen worden, wil de behandeling als effectief beoordeeld kunnen worden. Daarnaast zal een "hello-goodbye" effect of een welwillendheidseffect de antwoorden nog positief kunnen beïnvloeden. (Onder het "hello-goodbye" effect wordt verstaan de neiging tot het geven van een negatieve voorstelling van zaken aan het begin van een therapie en een te positieve voorstelling aan het einde ervan. Soudijn, 1976, p.68). Echter redenerend vanuit het standpunt dat hier sprake is van een groep patiënten met ernstige en vaak lang bestaande problemen, is een geringe verbetering wellicht toch waardevol, zeker wanneer deze verbetering blijvend is.

De komende evaluatieperiode zal kunnen blijken, of een bepaald effectiviteitscriterium kan worden gehanteerd, wanneer op reguliere basis, o.a. bij ontslag, evaluatieve gegevens verzameld worden.

8.3 DE PERIODE NOVEMBER 1976 - 1980.

Vanaf 15 november 1976 werd aan iedere patiënt bij ontslag een vragenlijst gegeven. Deze lijst kwam inhoudelijk overeen met de follow-up vragenlijst welke hiervoor beschreven is. Vragen over de periode na ontslag vervielen uiteraard. Bij enkele vragen werden de antwoord-alternatieven gewijzigd op basis van de opmerkingen van de invullers. Deze vragenlijst werd door drie follow-up lijsten gevolgd. Sinds 1977 werd ook de psychotherapeut gevraagd bij opname informatie te verstrekken, over de diagnose e.d., alsmede per 1-1-1978 bij ontslag over de resultaten van de behandeling. (zie hoofdstuk 6.3) In totaal werden in genoemde periode 310 patiënten ontslagen. De responsepercentages per formulier staan in tabel 8.1.

In deze tabel staan twee percentages vermeld: het percentage verzonden/uitgereikte formulieren en het percentage

Tabel 8.1 Aantallen verzonden en terugontvangen formulieren over de periode nov. 1976 tot 1981

formulier	totaal	verzonden		retour	
	N	N	%	N	%
statistiek	310	310	100	310	100
diagnose	292	292	100	292	100
ontslag patiënt	310	289	93	265	92
ontslag therapeut	216	207	96	207	100
1e follow-up pat. (3 mnd)	310	295	95	277	94
2e follow-up pat. (6 mnd)	310	276	89	244	88
3e follow-up pat. (12mnd)	310	265	85	225	85
1e follow-up ther.(3 mnd)	310	251	81	191	76
2e follow-up ther.(6 mnd)	310	211	68	149	71
3e follow-up ther.(12mnd)	310	159	51	129	81

retour ontvangen lijsten berekend naar het aantal verzonden lijsten. Dit laatste is in de meest letterlijke zin een antwoordpercentage. Het is van belang de motieven van het niet versturen en de motieven van het niet reageren te kennen. Vragenlijsten zijn niet verstrekt of verzonden, wanneer de patiënt slechts kort is opgenomen (minder dan 4 weken). Een dergelijke opname is dijkwijs een "drop out". Het lijkt niet opportuun dan nog tot een jaar na dato informatie te vragen. Dit betreft ca. 5% van alle patiënten. Een tweede reden om geen lijst te sturen is, wanneer de patiënt bij ontslag of later te kennen geeft geen lijst meer te willen invullen of bekend is dat hij geen lijst meer kan invullen (opname elders, overleden). Bij ontslag zeggen 13 patiënten (4% van in totaal 310 patiënten) op dat moment en later geen lijsten te willen invullen. Dit zijn doorgaans patiënten die zeer ontevreden vertrokken zijn.

Wanneer blijkt, dat een patiënt bij herhaling geen lijst retourneert, wordt verder geen informatie meer gevraagd. Dit doet zich vooral bij de derde follow-up vragenlijst voor. Bij de follow-up vragenlijst aan de behandelaar is het belangrijkste motief om geen lijst te sturen, wanneer er geen voortgezette hulpverlening (meer) is. Dit percentage stijgt uiteraard van de eerste naar de derde follow-up. Teneinde een beter zicht te krijgen op niet-verzend- en niet retourneerredenen worden deze sedert 1981 per vragenlijst en per patiënt bijgehouden. In paragraaf 8.4 worden deze vermeld.

Uit tabel 8.1 blijkt, dat de response bij de patiënten hoog is: 85-100%; bij de hulpverleners is dit lager: 71-100%. De respons neemt af naarmate meer tijd verstrijkt tussen ontslag- en follow-up moment. Geconcludeerd wordt, dat de responspercentages voldoende hoog zijn vergeleken met ander onderzoek om aan de uitslagen belang te hechten. De vraag in hoeverre de non-responders een selecte groep is, blijft

natuurlijk intrigeren. In hfst. 9.5 wordt nagegaan of deze groep minder effectief behandeld is. De resultaten van de evaluatieve vragenlijsten bij ontslag en follow-up worden hier per lijst weergegeven.

8.3.1 Ontslagvragenlijst patiënt.

Van 265 patiënten is een ontslagvragenlijst ontvangen. De antwoorden op de belangrijkste vragen met betrekking tot effect en satisfactie worden in vier jaren gepresenteerd. De 15 patiënten die eind 1976 zijn ontslagen, worden bij 1977 geteld.

Tabel 8.2 Behandelingseffect door patiënt beoordeeld (%).

Vraag: Wat gebeurde er met de problemen? Deze zijn:	77	78	79	80	T
1. Niet behandeld	3	2	5	4	3
2. Geheel/tam. goed verdwenen	66	58	49	43	55
3. Een beetje verdwenen	24	35	32	42	32
4. Hetzelfde/slechter geworden	7	6	14	11	10
Aantal patiënten	80	66	63	53	262

Tabel 8.2 laat zien, dat in totaal 55% van de patiënten aangeeft, dat de problemen (tamelijk) goed verbeterd zijn. Over de vier jaren is dit percentage gedaald van 66% naar 43%. Dit resulteert in een stijging van het percentage dat zegt "een beetje" verbeterd te zijn. Het gemiddelde behandelingsresultaat is kennelijk gedaald.

Twee andere effectvragen luiden: "Hoe voelt u zich nu met betrekking tot de problemen?" en "Hoe voelt u zich nu in vergelijking met het moment dat u opgenomen werd?". In tabel 8.3 en 8.4 staan de resultaten vermeld.

Tabel 8.3 Algemene effectvraag over status quo (%).

Vraag: Hoe voelt u zich nu?	77	78	79	80	T
1. Goed	5	6	9	4	6
2. Redelijk	76	71	59	68	69
3. Niet goed	14	17	27	22	19
4. Slecht	5	6	5	6	5

Tabel 8.4 Algemene effectvraag over verschil tussen opname en ontslag (%).

Vraag: Hoe voelt u zich nu vergeleken bij de opname?	77	78	79	80	T
1. Veel beter	56	48	46	30	46
2. Wel wat beter	33	41	41	54	41
3. Hetzelfde	5	5	5	6	5
4. Slechter	6	6	8	11	8

Uit de antwoorden in tabel 8.3 blijkt, dat 24% zegt dat het niet goed of slecht gaat en dat dit percentage over de vier jaren gestegen is van 19% naar ca. 30%. Het percentage dat zegt "goed", blijft constant. Een grote, maar licht dalende meerderheid zegt "redelijk". Bij de vraag naar het verschil tussen opname en ontslag (tabel 8.4) bestaat een overeenkomstige tendens: de meest positieve categorie daalt van 56% naar 30% ten voordele van het antwoord: "wel wat beter" dat stijgt van 33% naar 54%. De twee overige categorieën "hetzelfde" en "slechter" stijgen slechts gering. Uit deze antwoorden op de drie effectvragen kan geconcludeerd worden, dat het effect van de behandeling zoals dit door de patiënten beoordeeld wordt, dalende is.

Als kritiek kan vermeld worden, dat de antwoordcategorieën meer spreiding behoeven aan de positieve zijde van de schaal: de categorie "redelijk" en "wel wat beter" is te ruim (zie Hfst. 8.4). Naast deze directe beoordelingen van het effect van de behandeling, wordt ook gevraagd naar de tevredenheid over de behandeling zelf, de behandelaars en andere aspecten van de opname. Op de vraag of de juiste problemen behandeld zijn, zegt 82% dat dit inderdaad is gebeurd. Over de behandelingsmethoden is men minder eensluidend positief: 59% vindt dat deze goed zijn geweest, terwijl 41% zegt dat sommige beter hadden gekund, of dat (ook) andere nodig waren geweest. Het oordeel over de individuele psychotherapie wordt door 90% "goed" genoemd.

Ook de oordelen over de psychotherapeut zijn heel positief, zoals blijkt uit onderstaande cijfers:

- ik zag deze: heel veel (31%), tamelijk veel (59%);
- deze heeft me: heel goed (67%), tamelijk goed (26%) geholpen;
- het contact was over het algemeen: erg goed (55%), tamelijk goed (39%);
- het contact was over het algemeen: erg prettig (61%), tamelijk prettig (35%).

Deze resultaten zijn over de vier jaren heel stabiel, behalve de vraag over hoe goed de patiënt zich geholpen voelt. Het meest positieve antwoord daalt van 79% in 1978 naar 55% in 1980 en het antwoord "tamelijk goed" stijgt dien-

overeenkomstig van 14% naar 38%. Deze trend is overeenkomstig die welke gesignaleerd is bij de effectvragen.

De vragen met betrekking tot het ontslag leveren de volgende resultaten op:

- het ontslag was op tijd zegt 70%, (veel) te vroeg 25% en te laat 5%;
- de belangrijkste reden voor ontslag is, dat de problemen door de nazorg verder geholpen kunnen worden (59%), terwijl 13% zegt dat de problemen (geheel) over zijn. De overige 28% zegt, dat de problemen niet voldoende verbeterd zijn, maar dat de opnameduur te lang werd of dat ze onoplosbaar zijn (8%) of dat er om anderen redenen een einde aan de opname kwam (20%);
- bij 80% van alle patiënten is voortgezette hulpverlening geregeld, 11% krijgt geen hulp en bij 9% is bij ontslag nog geen hulp gevonden. Van degenen die geen nazorg krijgt, zegt de helft het niet nodig te hebben, de anderen het niet te willen.

Naast deze kwantitatieve gegevens, waren de invullers ook in de gelegenheid middels open vragen nadere toelichting te geven op diverse onderdelen van de opname. Dit betrof opmerkingen over het verblijf op Overwaal, het verloop van de behandeling en het ontslag. Een open vraag betrof de vragenlijst zelf: Wat vond u van het invullen van deze lijst. Wanneer deze werd ingevuld, was dit over het algemeen positief (69 keer). Kwalificaties als "moeilijk" e.d. werden 23 maal gegeven.

Aldus een overzicht van de belangrijkste antwoorden op de vragenlijst welke door de patiënt bij ontslag werd ingevuld. Samengevat blijkt, dat de satisfactie over de behandeling hoog en zeer stabiel is. De beoordeling van het effect van de behandeling is lager en vertoont een dalende tendens.

8.3.2 Ontslagvragenlijst therapeut.

Sinds 1978 vullen ook psychotherapeuten een ontslagvragenlijst in. Deze lijst is bij bepaalde vragen parallel aan die van de patiënt. Als belangrijkste reden voor ontslag wordt genoemd, dat de problemen geheel over zijn of dat de patiënt het verder zelf of met nazorg weer aan kan (57%). Bij 18% wordt de behandeling afgebroken, omdat Overwaal geen oplossing kon bieden en bij 24% is sprake van een afbreken van de behandeling door de patiënt. Voor 15% wordt geen nazorg geregeld, vooral omdat de patiënt dit niet wil of omdat dit beter geacht wordt. Het ontslag zelf wordt bij 59% als "op tijd" beoordeeld, bij 30% als "te vroeg" en bij 11% als "te laat". Het resultaat van de behandeling wordt beoordeeld door de mate van verschil tussen opname en ontslag aan te geven. In tabel 8.5a staan de antwoordpercentages. Hieruit blijkt, dat 42% als "veel beter" beoordeeld wordt, en 39% als "wel (wat) beter"; van 14% wordt gezegd dat patiënt zich "hetzelfde" voelt als bij opname en bij 4% is de situatie verslechterd. Er is sprake van een lichte daling van het oordeel in de

bovenste categorie en een stijging van het een na hoogste antwoordpercentage.

Tabel 8.5a Oordeel therapeut over het verschil tussen opname en ontslag over de jaren 1978, 1979 en 1980 (%).

	1978	1979	1980	Tot
1. veel beter	44	43	39	42
2. wel (wat) beter	36	39	44	39
3. hetzelfde	15	15	13	14
4. slechter	5	3	5	4
aantal patiënten	75	69	64	208

Tenslotte wordt gevraagd hoe de psychotherapeut verwacht dat het na ontslag zal gaan. Deze beoordeling is minder gunstig: bij 35% verwacht de therapeut dat het (misschien) niet goed zal gaan; "het gaat goed" luidt het oordeel over 12% van de patiënten en bij 52% zal het "met goede nazorg uiteindelijk wel lukken". In tabel 8.5b staan de resultaten per jaar vermeld.

Tabel 8.5b Verwachting van de therapeut over hoe het na ontslag zal gaan (%).

	1978	1979	1980	T
1. het gaat goed	16	12	8	12
2. met goede nazorg zal het uiteindelijk wel lukken	51	55	52	52
3. het zal misschien niet goed gaan	19	12	28	19
4. ik denk niet dat deze het goed zal maken.	13	22	12	16

Het verloop van de resultaten over de drie jaren is wisselend: een lichte daling van het antwoord: het gaat goed. De tweede antwoordcategorie heeft een betrekkelijk stabiel antwoordpercentage, terwijl de twee onderste antwoorden een sterk wisselend beeld geven. Worden deze samen genomen, dan blijkt er een stijging van de twee ongunstige verwachtingen van 32% naar 40%.

Samenvattend kan geconcludeerd worden dat de beoordelingen van de therapeut overeenstemmen met die van de patiënt: ruim

40% is veel verbeterd, 40% wel wat verbeterd en bijna 20% niet verbeterd of slechter geworden. Ook de trend van de resultaten is overeenkomstig: beide geven een geleidelijke vermindering van het behandelingsresultaat. Ook de verwachtingen van de therapeut over hoe het na ontslag zal gaan, worden minder positief.

8.3.3 Follow-up patiënt.

De follow-up vragenlijsten die aan de ex-patiënt na 3, 6 en 12 maanden worden gestuurd, betreffen vooral vragen over de status quo en beoordelingen van de behandeling op Overwaal en die na ontslag.

In de tabellen worden de resultaten per follow-up moment weergegeven over de opeenvolgende jaren, zoals dit elders ook steeds is gebeurd. In tabel 8.7, 8.8 en 8.9 staan per follow-up moment de gegevens over een belangrijke effectvraag: "Hoe voelt u zich nu met betrekking tot de problemen waarvoor u op Overwaal werd opgenomen?". Uit deze tabellen blijkt, dat de trend welke bij de ontslagvragenlijsten reeds bleek, zich bij de follow-up nog duidelijker manifesteert. Bij ontslag zei in 1977 81% zich goed/redelijk te voelen, hetgeen daalde naar 72% in 1980. Bij het eerste follow-up moment is dit resp. 60% in 1977 en 45% in 1980. Opmerkelijk is, dat ook hier de twee extreme categorieën nauwelijks veranderen, maar dat het percentage "redelijk" daalt en "niet goed" dienovereenkomstig stijgt. Bij het tweede en derde follow-up moment is deze trend eveneens zichtbaar. De daling bij de 1e follow-up is significant. Bij de 2e en 3e follow-up zijn deze verschillen, wellicht mede door de voortgezette hulpverlener, genivelleerd.

Tabel 8.6 1e Follow-up moment: Hoe voelt u zich nu m.b.t. de problemen waarvoor u werd opgenomen? (%)

	1977	1978	1979	1980	T
1. Goed	7	6	6	4	6
2. Redelijk	53	67	49	41	53
3. Niet goed	35	22	40	48	36
4. Slecht	5	6	5	7	5
aantal patiënten	83	69	67	56	275

F-ratio = 3.54; $p = 0.02$

Worden de follow-up resultaten vergeleken met die bij ontslag, dan blijkt er tussen ontslag en de eerste follow-up een opmerkelijk verval: het percentage patiënten dat zegt dat het

Tabel 8.7 2e Follow-up moment: Hoe voelt u zich nu m.b.t. de problemen waarvoor u werd opgenomen? (%).

	1977	1978	1979	1980	T
1. Goed	6	8	5	8	7
2. Redelijk	63	66	57	45	58
3. Niet goed	26	19	29	37	27
4. Slecht	6	7	9	10	8
aantal patiënten	70	59	63	51	243

F-ratio = 2.12; p = 0.1

Tabel 8.8 3e Follow-up moment: Hoe voelt u zich nu m.b.t. de problemen waarvoor u werd opgenomen? (%).

	1977	1978	1979	1980	T
1. Goed	15	18	14	9	14
2. Redelijk	53	57	58	50	55
3. Niet goed	29	17	23	41	27
4. Slecht	3	8	5	0	4
aantal patiënten	62	60	57	46	225

F-ratio = 1.18; p = 0.3

niet goed of slecht is bij ontslag, is ca. 25%. Bij de eerste follow-up is dit gestegen tot 41%, waarna dit weer geleidelijk daalt naar 35% en 31% bij de tweede en derde follow-up. In tabel 8.9 staan deze gegevens vermeld, en is in feite een samenvatting van de tabellen 8.3, 8.6 t/m 8.8 t.a.v. de twee onderste antwoordcategorieën.

Tabel 8.9 Ontslag en follow-up gegevens. Van de vraag: "Hoe voelt u zich nu? worden de twee laagste antwoordcategorieën "niet goed" en "slecht" vermeld. (in %).

	1977	1978	1979	1980	tot
ontslag	19	23	32	28	24
1e follow-up	40	28	45	55	41
2e follow-up	32	26	38	47	35
3e follow-up	32	25	28	41	31

Geconcludeerd kan worden, dat het behandelingseffect gedurende de beschreven periode volgens deze follow-up gegevens, geleidelijk vermindert, hetgeen met name bij het eerste follow-up moment blijkt. (Het jaar 1978 wijkt van deze trend af.) Het grote verval na 3 maanden blijkt bij de verdere follow-up weer te stijgen, maar bereikt niet het niveau van bij ontslag. In hoeverre het ontslagresultaat vertekend wordt door antwoordtendenzen is niet aantoonbaar, maar wel aannemelijk.

De overige vragen van de drie follow-up lijsten worden hier niet in extenso beschreven. Volstaan wordt met een samenvatting van de resultaten.

Op de vraag of er verbeteringen zijn opgetreden sinds ontslag, zegt bij de eerste follow-up 52% dat ze zich (veel) beter voelen, 23% zegt zich hetzelfde te voelen en 25% voelt zich slechter dan bij ontslag. Bij de tweede follow-up zijn deze cijfers resp. 61%, 18% en 21% en bij de derde follow-up: 70%, 13% en 17%. Hieruit blijkt, dat de meerderheid van de patiënten na ontslag een verbetering aangeven, welke toeneemt naarmate de follow-up periode langer wordt. Er blijkt een sterk verband te bestaan tussen deze vraag en de vraag naar de status quo: wanneer patiënten zeggen dat het goed of redelijk gaat, geven ze ook aan dat ze verbeterd zijn sinds ontslag; gaat het "niet goed" of "slecht", dan melden ze een verslechtering (χ^2 op alle drie de follow-up momenten is significant, $p < .001$).

Het niveau van functioneren blijkt eveneens geleidelijk te verbeteren. Bij de 1e follow-up zegt 45% dat men (met moeite) alles weer kan doen, welke percentage stijgt naar 63% bij de derde follow-up. De twee laagste antwoordcategorieën en "ik kan weinig/eigenlijk niets doen" worden constant door ca. 15-20% aangekruist.

Het percentage patiënten dat geen hulpverlening (meer) heeft gekregen, stijgt van 16% naar 37% bij de derde follow-up. Dit is echter niet zonder meer een positieve ontwikkeling, slechts een kwart van hen zegt geen hulp nodig te hebben, omdat de problemen over zijn. Bij de overigen is de hulpverlening afgebroken door de patiënt (50%) of door de hulpverlener (25%). Op de vraag of men op dit moment hulp wenst, zegt een stijgend percentage geen hulp nodig te hebben (van 15% naar 35%), een constant percentage van ca. 15% zegt geen hulp te hebben maar deze wel te wensen en ca. 8-15% wenst graag andere hulp. De overige 40-50% is tevreden met hun hulpverlener.

Bij de eerste follow-up werd ook nog gevraagd naar ervaring met betrekking tot afwijzingen, omdat men opgenomen is geweest. Dit zegt 14% wel eens ervaren te hebben.

Tenslotte werd ook gevraagd naar het oordeel achteraf over de opname op Overwaal en de duur ervan:

- zeer tevreden zegt 42% (46% van 1977 en 36% in 1980).
- tamelijk tevreden zegt 36% (stijgt van 34% naar 40%).
- wisselend of weet ik niet, zegt 16% (van 14% naar 18%).
- tamelijk of zeer ontevreden, zegt constant ca. 7%.

Over de duur van de opname zegt 57% dat deze juist goed was, ruim 30% vond deze te kort en ca. 10% te lang. Deze cijfers wijken licht af van hetgeen bij ontslag gezegd was, toen 70% meende dat het tijdstip van ontslag "juist goed" was.

Hiermee zijn de belangrijkste resultaten van de drie follow-up lijsten van de patiënt beschreven. Er blijkt een belangrijke verslechtering aangegeven te worden bij het eerste follow-up moment. Bij tal van vragen treedt er na verloop van tijd weer een verbetering op. Voortgezette hulpverlening blijft bij de overgrote meerderheid gedurende de hele follow-up periode bestaan of gewenst. Een klein gedeelte meldt, dat de problemen over zijn. De gesignaleerde verslechtering bij de eerste follow-up wordt wellicht mede veroorzaakt door een vertekening van de antwoorden bij ontslag door een "hello-goodbye" effect.

8.3.4 Follow-up nazorg.

Tegelijk met de patiënt ontving ook zijn hulpverlener een vragenlijst, uiteraard met zijn instemming. Bij de beoordeling van de resultaten moet rekening gehouden worden met het feit, dat alleen over die patiënten informatie verkregen wordt die nog hulpverlening ontvangen. Uit het gegeven dat wanneer geen hulpverlener bezocht wordt, dit slechts bij 25% als reden geeft dat de problemen over zijn, hoeft deze selectieve informatie niet noodzakelijkerwijze te betekenen, dat dit een vertekening is, in die zin dat alleen van patiënten die minder verbeteren, informatie van zijn behandelaar na ontslag verkregen is.

Ten aanzien van de effectvragen treedt hier dezelfde trend op: een stijgend percentage zegt dat het "niet zo goed" of "slecht" gaat met de patiënt bij de eerste follow up: van 35% naar 54%, hetgeen geheel overeenstemt met de antwoorden van de patiënten (zie tabel 8.6).

Er is een verschil in vraagstelling naar de vorderingen sinds ontslag tussen de eerste follow-up lijst en die van de tweede en derde, waardoor deze niet zonder meer op één lijn zijn te stellen. Bij de 1e follow-up meldt 32% van de nazorgers dat de problematiek hetzelfde is en 20% dat deze verslechtert (stijgt van 15 naar 23%). Bij de 2e en 3e follow-up lijst zijn de antwoordcategorieën een 7-punts schaal die loopt van: 1 = cliënt maakt sterke vorderingen tot 7 = cliënt gaat sterk achteruit. De modus ligt bij punt 3, dat door 51% resp. 42% bij de 2e en 3e follow-up wordt aangekruist. Uit de verdeling over de antwoorden blijkt een verbetering op te treden in de tijd, het gemiddelde oordeel gaat van 3,1 naar 2,8. Op de overige vragen blijkt weinig verschil tussen het 2e en 3e follow-up moment: het niveau van functioneren wordt bij ruim 40% normaal genoemd, 21% kan met hulp/beschermd normaal functioneren, 20% wisselend en ca. 15% functioneert (hoofdzakelijk) niet. Ook hier is de trend dalend gedurende de vier jaren.

Bij de eerste follow-up lijst werd ook gevraagd naar oordelen over de doorverwijzing, de informatie over de client (uitschrijfbrief) en de behandeling op Overwaal.

Samengevat zijn deze resultaten.

- Tijdstip van contact leggen met de voortgezette hulpverlener: te vroeg 13%, op tijd: 61%, te laat: 26%.
- Tevredenheid regelen nazorg: zeer tevreden: 24%, tamelijk tevreden: 45%, weet niet 15%, ontevreden: 16%.
- De toegezonden informatie werd (zeer) uitvoerig gevonden (75%), (helemaal) juist (73%) en men zegt er (heel) veel aan te hebben gehad (58%).
- De verwijzing was (zeer) terecht (80%) en de therapie sloot (erg) goed aan (59%)

De middencategorie wordt steeds door ca. 15% aangekruist. Negatieve oordelen komen slechts zelden voor. Ook bij deze beoordelingen is er overigens sprake van een dalende tendens.

Het oordeel over de behandeling op Overwaal staat in tabel 8.10 vermeld. Hieruit blijkt, dat ruim 50% de opname zinvol vindt en 24% zegt dat "opname wel zin had"

Het meest gunstige oordeel daalt van 63% naar 41%, terwijl de twee ongunstige oordelen samen stijgen van 19% naar 28%.

Tabel 8.10 Oordeel over de opname door de voortgezette hulpverlener bij de eerste follow-up (%).

	1977	1978	1979	1980	T
1 Goed, was zinvol	63	54	57	41	54
2. Redelijk, had wel zin	19	24	23	31	24
3. Matig, heeft niet veel effect gehad	15	22	18	28	20
4 Slecht, heeft client geen goed gedaan	4	0	2	0	2
aantal patienten	48	50	40	39	177

8.4 CONCLUSIE

In het voorafgaande is een melé van gegevens gepresenteerd op univariaat of bivariaat niveau. Hieruit blijkt een geleidelijke daling van het behandelingsresultaat. Deze daling wordt consistent door beide typen informanten aangegeven. Dit is als zodanig een product dat een dergelijk project kan opleveren. Deze resultaten zijn periodiek aan de instelling gerapporteerd om datgene er mee te doen, wat men wenselijk achtte.

Vanzelfsprekend dringen zich ook vragen op die het descriptieve niveau overstijgen

- Wat is de samenhang tussen deze gegevens, welke van verschillende informanten. patient, psychotherapeut en voortgezette hulpverlener zijn verkregen, op meerdere meetmomenten: ontslag en 3 follow-up metingen en van zeer uiteenlopende inhoud: evaluatie van effecten, beoordelingen van status quo, procedures en satisfactie.
 - Welk verband bestaat er tussen de zo veelvuldig gesignaleerde trend van een vermindering van het behandelingsresultaat met andere variabelen, zoals demografische gegevens, problematiek, psychotherapeut, e.d.
- Deze vragen zullen in hoofdstuk 9 beantwoord worden.

8.5 DE PERIODE 1981-1985.

In de loop van 1980 werden de ontslag- en follow-up vragenlijsten herzien. (zie hoofdstuk 6). Vanaf januari 1981 vulden alle patienten deze nieuwe lijsten in. Bij de invoer hiervan werd ook bijgehouden of er al dan niet een lijst uitgereikt of verzonden was en zo niet, met welke redenen. Ook als een lijst niet werd geretourneerd, werd getracht na te gaan waarom dit het geval was. Na een schriftelijk rappel volgde zo mogelijk nog een telefonisch contact om het motief van het niet reageren te achterhalen. Deze inspanningen deden de responspercentages stijgen. Dikwijls bleek dat het feit dat nog niet gereageerd was, gelegen te zijn in vergeetachtigheid. Actief weigeren om aan deze evaluatie mee te doen - bij ontslag of later - kwam weinig voor, in totaal ca. 5%. De belangrijkste redenen dat geen lijst verzonden werd, waren: heropname in Overwaal, te korte opname (< 4 weken) of geen voortgezette hulpverlener. In enkele gevallen werd geen lijst verzonden, omdat de patient overleden was of niet in staat (bijv. psychotisch) te reageren.

In tabel 8.11 staan de respons-percentages vermeld. Onder de categorie "niet verzonden" worden niet gerekend patienten die bij ontslag reeds te kennen geven niet te willen meedoen. Deze zijn geteld tot de non-responders. Uit tabel 8.11 blijkt, dat het respons-percentage daalt tot ruim 70% wanneer de follow-up periode langer wordt. Overigens kunnen deze percentages bevredigend worden genoemd, zeker voor wat betreft de vragenlijsten die bij ontslag en 1e follow-up zijn ingevuld. Op deze momenten wordt meer dan 90% van de lijsten geretourneerd door de patienten en hun hulpverleners.

In de loop van 1985 is de invoer van gegevens gestopt. Over dit jaar is slechts een gedeelte van de ontslagvragenlijsten van de patient en de therapeut beschikbaar. Deze gegevens blijven buiten beschouwing in tabel 8.11. Voor zover mogelijk worden de resultaten over 1985 wel vermeld in paragraaf 8.4.2 en 8.4.3.

Zoals in de vorige paragraaf worden per vragenlijst de belangrijkste resultaten zonodig bivariaat weergegeven, wanneer een mogelijke trend gedurende de vier of vijf beschreven

Tabel 8.11 Aantallen verzonden en terugontvangen lijsten over 1981-1985, met motieven waarom geen lijsten zijn verzonden

lijst	tot.	---niet-verzonden---				verz. N	ret. %
		her.	kort	ov.	%		
statistiek	333	-	-	-	-	333	100
diagnose	333	6	14	-	6	313	100
opname	333	22	10	-	10	277	91
ontslag ther.	333	25	4	-	9	302	99
ontslag pat.	333	20	24	-	13	257	88
f.u.pat. 1	333	18	24	9	17	267	94
f.u.pat. 2	333	22	24	8	16	248	88
f.u.pat. 3	331	30	24	7	18	211	72
f.u.nazorg 1	290*	19	24	5	17	223	91
f.u.nazorg 2	258*	22	24	8	21	176	84
f.u.nazorg 3	234*	28	24	4	24	142	75

* vermeld is het aantal ontslagen patiënten dat ten tijde van het follow-up moment een voortgezette hulpverlener had.
tot. = totaal; her. = heropname; kort = te korte opname
ov. = overige redenen; verz. = verzonden; ret. = retour

jaren zichtbaar is. Buiten beschouwing blijven het aanvraag- en intake-formulier, omdat deze slechts registratieve informatie bevatten. Verschillen tussen de periode 1976 t/m 1980 en de hier beschreven periode 1981-1985 komen in paragraaf 8.5 aan de orde.

8.5.1 Opnamevragenlijst patient.

Opmerkelijk is, dat 70% van de patiënten bij opname "nog niet zo veel" of "eigenlijk niets" over Overwaal weten. Het merendeel is naar Overwaal gegaan omdat zij dit helemaal (53%) of hoofdzakelijk (30%) zelf wilden. Men ziet heel/tamelijk erg tegen deze opname op (61%) en wordt heel/tamelijk erg (64%) belemmerd in het dagelijks leven door problemen. Van de maximaal drie door de patiënt geformuleerde problemen wordt ook gevraagd wat de verwachting is ten aanzien van de verbetering bij ontslag. Tussen de genoemde problemen blijkt nauwelijks verschil te bestaan ten aanzien van deze verwachting. Ca. 16% verwacht dat de problemen geheel zullen verbeteren, 45% "voor het grootste gedeelte" en 29% zegt te verwachten dat ze "min of meer" zullen verbeteren. "Weinig" of "niet" zegt in totaal 10%.

8.5.2 Ontslagvragenlijst patiënt.

De eerste evaluatie van het behandelingsresultaat zijn de door de patiënt bij opname beschreven problemen/doelen. Deze zijn uitgetypt en bij de ontslagvragenlijst gevoegd. Gevraagd wordt per doel aan te geven of deze behandeld zijn en of ze verbeterd zijn. Van het eerste, en doorgaans ook het belangrijkste probleem zegt 78% dat dit ook behandeld is. De verbetering kan op een zespuntsschaal worden aangegeven: van geheel verbeterd tot slechter geworden. In tabel 8.12 staan de antwoordpercentages op deze vraag vermeld.

Tabel 8.12 In hoeverre is het eerst vermelde probleem bij ontslag verbeterd?

	1981	1982	1983	1984	1985	T
1. geheel	11	12	5	5	3	7
2. het grootste gedeelte	27	32	20	49	31	33
3. min of meer	36	27	39	27	27	31
4. weinig	16	19	12	9	22	15
5,6. niet/verslechterd	10	10	24	10	17	13
aantal patiënten	62	59	59	75	64	319

Uit deze tabel blijkt, dat 40% zegt dat de problemen voor het grootste gedeelte of geheel verbeterd zijn, 31% zegt min of meer, terwijl 28% weinig, geen verbetering of zelfs een verslechtering aangeeft.

Het eerder genoemde probleem van de hoogte van het criterium doet zich nu gelden: welke antwoordcategorieën representeren een niet-effectieve behandeling? In vergelijking met ander effectonderzoek zoals Malan (1973) en Sloane e.a. (1975), die dezelfde meetschaal hanteren, is gekozen voor een samenvoeging van de onderste drie categorieën: weinig verbeterd, niet verbeterd en verslechterd. Een dergelijke tweedeling lijkt arbitrair, maar gelet op de ernst van de behandelde patiëntengroep, de intensiteit van de behandeling en de mogelijkheid van een positieve antwoordtendentie bij ontslag, lijkt deze norm aannemelijk. Bij interne rapportages op Overvaal is deze ook gehanteerd en geaccepteerd. Bij de volgende beschrijvingen van de evaluatie tijdens de follow-up periode zal van deze norm eveneens worden uitgegaan.

Worden dus de drie onderste antwoordcategorieën samengenomen, dan geeft dit voor de opeenvolgende jaren als resultaat resp. 26%, 28%, 36%, 19% en 39%. Hieruit blijkt, dat het jaar 1984 sterk afwijkt van de trend dat het behandelingsresultaat geleidelijk minder wordt, thans volgens het oordeel van de patiënten. Bij de overige problemen zegt resp. 28% en 37% van

de patiënten dat dit bij ontslag weinig/niet verbeterd of verslechterd is. Met name bij het derde probleem bestaat er een grote schommeling tussen de vijf jaren; voor de onderste drie categorieën zijn de percentages: 23, 43, 45, 21 en 24%. Nog slechts de helft van de patiënten geeft een derde probleem op, waardoor de aantallen per jaar klein worden, hetgeen de kans op toevallige veranderingen doet toenemen. Ruim 50% zegt bij ontslag, dat gedurende de opname nog andere problemen gesignaleerd zijn. De antwoordpercentages voor deze problemen wijken nauwelijks af van die welke wel bij opname zijn genoemd.

Bij opname wordt per geformuleerd probleem ook aangegeven in hoeverre de patiënt verwacht dat dit bij ontslag zal zijn verbeterd. Bij ontslag blijken de resultaten hierbij achter te blijven. Wanneer de verwachtingen t.a.v. probleem 1 en de gemelde verbeteringspercentages met elkaar vergeleken worden dan geeft dit onderstaande resultaten:

antwoordcategorie	1	2	3	4	5+6
verwachting	15	49	28	6	2
resultaat	7	33	31	15	13

Dit verschil is statistisch significant ($\chi^2 = 32.6$ $p < .001$).

De resultaten op de globale effectvragen: Hoe voel je je nu? en Wat is het verschil tussen opname en ontslag? staan in tabel 8.13 en 8.14 vermeld.

Tabel 8.13 Globale effectvraag: Hoe voel je je nu? (%).

	1981	1982	1983	1984	1985	T
1. uitstekend	-	7	3	4	1	3
2. goed	29	38	17	33	37	31
3. gaat wel	47	37	42	49	30	41
4. niet zo goed	21	12	22	9	22	17
5. slecht	3	7	17	4	9	8
aantal patiënten	62	60	60	75	64	321

Deze vragen geven een opmerkelijk wisselend beeld. Met name 1983 laat een opvallend ongunstig antwoordpatroon zien. Worden in tabel 8.13 en 8.14 de twee onderste categorieën samengenomen, dan zegt resp. 24, 19, 39, 13 en 31% dat het bij ontslag niet zo goed of slecht gaat en 18, 17, 23, 9 en 17% zegt dat het hetzelfde of slechter gaat dan bij opname. Het eerder genoemde probleem waar het criterium gelegd moet worden voor een voldoende mate van verbetering blijft ook hier bestaan: de middelste antwoordcategorie bevat 30 resp. 41% van alle antwoorden. Deze verschillend typen effectvragen geven een wisselend beeld te zien: uit de probleem-

Tabel 8.14 Wat is het verschil tussen opname en ontslag.

	1981	1982	1983	1984	1985	T
1. veel beter	21	27	17	24	14	21
2. beter	26	22	28	43	37	32
3. wel wat beter	36	34	32	24	25	30
4. hetzelfde	11	10	3	4	8	7
5. slechter	6	7	20	5	9	10

evaluaties blijkt, dat 40% een goed resultaat aangeeft, 30% een matig resultaat en 30% weinig of geen resultaat. Bij de globale vragen is dit voor wat betreft de status quo bij ontslag 35 - 40 - 25 en qua verschil tussen opname en ontslag: 50 - 30 - 20.

Ook tussen de 5 jaren bestaan verschillen: bij de doelevaluatie kan een trend van verminderd resultaat gesignaleerd worden met een gunstige uitzondering van 1984. Bij de globale effectvragen is een dergelijk patroon minder herkenbaar, 1984 laat wederom een gunstig resultaat zien, maar met name 1983 laat een betrekkelijk ongunstige evaluatie zien: 20% geeft aan dat het slechter gaat dan bij opname.

De tevredenheid over de behandeling werd o.a. gemeten door een vertalen van de Client Satisfaction Questionnaire (zie hfst. 4). De resultaten van deze 8 vragen tellende satisfactieschaal staan in tabel 8.15 vermeld.

Tabel 8.15 Satisfactieschaal. Per vraag worden de twee positieve antwoordpercentages, hun som en de gemiddelde score (5 is max. positief) vermeld.

Vraag inhoud:	++	+	som	gem.
anderen adviseren	61	27	88	4,46
hoeveelheid hulp	52	36	88	4,32
algemene tevredenheid	40	47	87	4,21
behandeling gekregen die nodig was	40	42	82	4,19
eventueel weer terugkomen	43	28	71	4,01
kwaliteit van de behandeling	26	54	80	3,99
soort behandeling	34	37	71	3,91
beter met problemen omgaan	23	49	72	3,80

gem. = gemiddelde; vragen zijn volgens het gemiddelde gerangschikt. (N = 319)

Deze acht satisfactievragen geven een zeer gunstig beeld: patiënten zijn zeer tevreden over de behandeling. Opmerkelijk is het verschil tussen de algemene tevredenheid (87%) en het antwoord op de vraag of men beter met de problemen kan omgaan (72%). Of men al dan niet zonodig weer zou willen terugkomen, refereert wellicht ook aan andere factoren dan effect of satisfactie. Tussen de eerste vier jaren bestaan geen grote of trendmatige verschillen. In 1985 blijkt echter een duidelijke daling op te treden bij alle satisfactievragen van ca. 15-20%. Aangezien dit geen volledig jaar is, kunnen hier geen conclusies aan verbonden worden.

Gezien het feit, dat de verdeling van de antwoorden zeer scheef is naar het positieve deel van de schaal, zijn in tabel 8.15 beide positieve antwoorden vermeld. Hieruit blijkt, dat niet altijd het meest positieve antwoord het frequentst is aangegeven. Als er sprake is van een antwoordtendentie door een "hello-goodbye" effect, dan kan wellicht een klein verschil reeds van betekenis zijn. De rangvolgorde op basis van gemiddelde score zou dan meer kunnen aangeven dan het verschil van 0,66 tussen de eerste en achtste vraag zou doen vermoeden. Opmerkelijk is, dat meer globale tevredenheidsvragen (vraag 1, 2 en 3) hoger scoren dan de specifiekere vragen als soort, kwaliteit en effect van de behandeling. Overigens kan geconcludeerd worden dat de in hfst. 4 gegeven suggestie om de antwoordmogelijkheden uit te breiden, weinig effect heeft gehad. Een eventuele uitbreiding dient nog meer aan het positieve eind van de schaal gezocht te worden.

De overige vragen welke bij ontslag gesteld worden, betreffen oordelen over psycho- en sociotherapeut, de medepatiënten en de ontslagprocedure. De beoordeling van de psychotherapeut respectievelijk sociotherapeut is als volgt. Men zegt (de antwoordpercentages voor de psycho- en sociotherapeut staan tussen haakjes):

- voldoende gesprekken te hebben gehad (88%-58%)
- dat deze (bijna) altijd bereikbaar was (95%-90%)
- dat de aanpak (tamelijk) goed aansloot (86%-72%)
- (zeer) tevreden over hem/haar te zijn (89%-73%)
- achteraf (misschien) geen andere gehad te willen hebben (77%-70%).

De beoordelingen van de sociotherapeut zijn iets lager dan die van de psychotherapeut; gezien de functionele verschillen is dit begrijpelijk. Geconcludeerd kan worden dat de oordelen over de behandelaars over het geheel genomen (zeer) goed zijn.

De contacten met de medepatiënten worden minder gunstig beoordeeld:

- ik had tamelijk/heel veel contact: 51%
- ze hebben me tamelijk/heel goed geholpen: 46%
- het contact was tamelijk/erg goed: 51%
- het contact was tamelijk/erg prettig: 71%
- ik voelde me er (tamelijk) goed bij horen: 53%.

- het woonmilieu vond men tamelijk/erg gezellig (37%), 41% zegt "ging wel". (Er wordt overigens nieuwbouw gepland.)
- de leefregels vond men goed (90%).
- de begeleiding van het woonmilieu door de sociotherapeuten wordt slechts door 35% (heel) goed genoemd, terwijl 36% zegt: "matig" of "niets van gemerkt". Aangezien intern verschil van mening bestaat of sociotherapeuten ook als zodanig dienen te functioneren, is dit resultaat begrijpelijk.

De vragen over de ontslagprocedure betreffen het tijdstip en de reden voor ontslag. In totaal zegt 69% op tijd ontslagen te zijn; 27% vindt dit echter (veel) te vroeg. De reden voor ontslag was bij 61% van alle patiënten dat men met ambulante hulp weer verder geholpen kan worden, terwijl 16% zegt het verder zelfs zonder hulp aan te kunnen. Gelet op de doelstelling van Overwaal stemt het ontslag van 77% van de patiënten hiermee overeen, namelijk er is (zo nodig) ambulante hulp gerealiseerd.

Tenslotte wordt er een vraag gesteld over ervaren stigmatisering; 25% zegt weinig of geen begrip ondervonden te hebben in hun omgeving t.a.v. hun opname, bij 67% toonde men wel begrip, terwijl 8% geen reactie ontving.

Hiermee zijn de belangrijkste vragen beschreven die bij ontslag aan de patiënt zijn voorgelegd. Geconcludeerd kan worden, dat de tevredenheid over tal van aspecten van de behandeling (zeer) positief is, maar dat de beoordeling van het effect minder gunstig is en een dalende trend vertoont, met name op het niveau van de concrete doelstellingen.

8.5.3 Ontslagvragenlijst therapeut.

De vragenlijst die de therapeut bij ontslag invult, informeert enerzijds naar het behandelingsresultaat en hiermee samenhangende aspecten en anderzijds naar de ontslagprocedure, in feite controlerend of deze volgens de afspraken is uitgevoerd. Ten derde worden achteraf de intakecriteria beoordeeld.

Tabel 8.16 Effect van de behandeling, beoordeeld door de psychotherapeut (%).

antwoord: Pt. maakte:	81	82	83	84	85	T
goede vorderingen	34	31	25	35	34	32
redelijke vorderingen	29	31	41	37	39	35
weinig vorderingen	29	32	22	25	24	27
geen vorderingen	8	6	12	3	3	6
aantal patiënten	79	68	73	79	59	358

Tabel 8.17 Zinvolheid van de opname, beoordeeld door de psychotherapeut (%).

antwoord	1981	1982	1983	1984	1985	T
erg zinvol	44	34	29	53	51	42
wel zinvol	46	52	57	36	41	46
nauwelijks zinvol	6	13	7	11	7	9
niet zinvol/nadelig	4	1	7	0	2	3
aantal patiënten	79	68	75	80	59	361

Het effect van de behandeling wordt "goed" of "redelijk" genoemd door 67% van de therapeuten (zie tabel 8.16). Opvallend is de dalende lijn van het meest gunstige antwoord van 34% naar 25% om in 1984 weer te stijgen. Het op een na gunstigste antwoord vertoont een complementair verloop. De zinvolheid van de opname wordt hoger ingeschat, zoals uit tabel 8.17 blijkt: 87% vindt de opname erg/wel zinvol geweest. Hier is dezelfde trend zichtbaar als beschreven is bij het effect van de behandeling. In 1985 wordt de behandeling zelfs door 92% zinvol genoemd. Op de vraag of de problemen in voldoende mate zijn opgelost, zijn de oordelen minder positief: op een vijf puntsschaal geeft 17% een 5, het meest positieve antwoord. Vervolgens zijn de percentages: 24%, 23%, 15%, terwijl 21% het meest ontkennende antwoord geeft. De "ik twijfel" en ontkennende antwoordcategorieën krijgen samen 59% van de antwoorden.

De motivatie van de patiënt wordt constant als goed (47%) of redelijk (27%) beoordeeld. Over de relatie tussen therapeut en patiënt is de eerste (zeer) tevreden (42%) of tamelijk tevreden (47%). Andere aspecten van deze relatie worden eveneens (zeer) goed beoordeeld. De omgang was heel/tamelijk prettig (70%), het contact was heel/tamelijk goed (75%), de problematiek was heel/tamelijk boeiend (77%) en de therapeut voelde zich bij de patiënt en zijn problematiek heel/tamelijk betrokken (86%). Er bestaan nauwelijks verschillen tussen de jaren.

De verwachting over hoe het na ontslag met de patiënt zal gaan, is niet al te hoog gespannen. Uit tabel 8.18 blijkt dat slechts 6% meent, dat het zeker goed zal gaan, 30% zegt "het zal wel goed gaan", en 40% twijfelt. Over de vier jaren is er sprake van een duidelijk verloop: een daling van de positieve verwachting met een stijging van het "ik twijfel"-antwoord.

Bij ontslag heeft 78% een externe voortgezette hulpverlener. Als er geen hulp geregeld is, wordt dit niet nodig gevonden (39%) of door de patiënt niet gewild (17%). Bij de overige 44% was het niet mogelijk voortgezette hulp te verzorgen o.a. door een vroegtijdig vertrek van de patiënt. De interne afspraken over het regelen van nazorg zijn, dat min-

Tabel 8.18 Verwachting over hoe het na ontslag zal gaan, beoordeeld door de psychotherapeut (%).

antwoord	1981	1982	1983	1984	1985	T
Het gaat zeker goed	13	10	1	4	0	6
Het zal wel goed gaan	33	26	24	34	34	30
Ik twijfel	25	49	49	36	43	40
Ik denkt niet dat het goed zal gaan	18	6	13	19	9	14
Het gaat hoogst- waarschijnlijk niet goed	11	9	12	7	14	10
aantal patiënten	79	68	75	80	58	302

stens 3 weken voor ontslag contact met de hulpverlener moet zijn opgenomen, dat patiënt en/of therapeut hier een gesprek mee moet hebben gehad en er binnen twee weken een uitschrijfbrief moet zijn verstuurd. Deze afspraken worden, volgens de therapeut, bij ca. 80% van de patiënten nagekomen.

Ten aanzien van de beeindiging van de behandeling wordt door 34% gemeld, dat deze moeilijk is verlopen of zelfs is afgebroken. Dienovereenkomstig vindt men in 66% van de gevallen het ontslag "op tijd". Geconcludeerd kan worden, dat de ontslagprocedure doorgaans goed verloopt: op tijd wordt voortgezette hulp geregeld als dit is aangewezen en ook andere afspraken worden goed nagekomen.

Tenslotte worden de intakecriteria beoordeeld. Door de intaker worden de 14 criteria gewaardeerd op een vijfpuntschaal. Wanneer 'n één of twee wordt gegeven, betekent dit doorgaans een contra-indicatie en wordt de patiënt afgewezen. Dit impliceert dat de spreiding van scores van opgenomen patiënten gering is en varieert van 3 tot 5. De psychotherapeuten geven doorgaans bij ca. 5% wel de score 1 of 2. Een uitzondering hierop vormen de criteria: behandelingsduur met 21% beoordelingen lager dan drie, persoonlijkheidsstoornis (16%), motivatie (12%), coöperatie eigen milieu (12%) en externalisatie (10%). Deze resultaten zijn niet eenduidig te interpreteren. Het is mogelijk dat de beoordelingen mede worden bepaald door het behandelingsresultaat: wanneer het resultaat minder is dan gewenst, kan hiervoor een verklaring worden gezocht in een contra-indicatie die bij opname niet juist is ingeschat. Beoordeling van deze indicaties bij ontslag is derhalve geen juiste manier om de intake te evalueren. Dit zou veel eerder - bijv. na een maand - moeten geschieden, wanneer het behandelingsresultaat nog niet bekend is. Deze procedure is dan ook sinds 1982 ingevoerd. De resultaten hiervan zijn nog niet bekend.

Hiermee wordt de beschrijving van de therapeut-ontslagvragenlijst beëindigd. In tegenstelling tot de beoordelingen van de patiënt is bij het oordeel van de therapeut van de

effectiviteit van de behandeling geen dalende trend waarneembaar. Deze blijven onverminderd positief; ook de andere beoordelingen zijn gunstig. Anders is dit evenwel voor de prognose over hoe het na ontslag met de patiënt zal gaan. Hier is een sterke toename van de "ik twijfel" en de verwachting dat "het niet goed zal gaan"-antwoorden, welke niet met deze globale beoordelingen overeenstemmen. Deze prognose komt wel overeen met het patiëntenoordeel.

8.5.4 Follow-up vragenlijsten patiënt.

Het belangrijkste onderdeel van de follow-up vragenlijsten zijn de vragen over de status quo en het effect van de behandeling. Zoals bij de ontslagvragenlijst, worden de problemen zoals door de patiënt bij opname of ontslag genoemd, bij de vragenlijsten gevoegd en gevraagd wordt om per probleem aan te geven of dit is verbeterd. Aan de patiënt worden aldus één tot maximaal vijf problemen voorgelegd. Aangezien de evaluatiescores tussen deze vijf problemen weinig verschillen en het aantal patiënten dat meer dan één probleem noemt, snel afneemt, wordt hier volstaan met het eerste probleem, dat doorgaans ook het belangrijkste is.

Tabel 8.19 In hoeverre is het eerst vermelde probleem bij de eerste follow-up verbeterd (%).

Is dit probleem verbeterd?	1981	1982	1983	1984	T
1. geheel	9	8	6	3	6
2. voor het grootste gedeelte	23	18	19	43	27
3. min of meer	31	32	30	25	29
4. weinig	25	13	14	15	17
5,6. niet/verslechterd	12	29	31	14	21
aantal patiënten	68	60	64	75	267

Tabel 8.19 laat zien, dat de bij ontslag gesignaleerde trend ook uit de resultaten bij de follow-up blijkt: een stijging van het percentage dat zegt dat het probleem weinig/niet verbeterd c.q. verslechterd is, van 37% naar 45% in 1983, en daalt weer naar 29% in 1984.

Bij de twee volgende follow-up momenten treedt er weer een licht herstel op, zonder evenwel het niveau zoals dat bij ontslag aangegeven werd, te bereiken. Conform tabel 8.9 worden in tabel 8.20 de resultaten van de vier meetmomenten bij elkaar gezet; volstaan wordt met de gesommeerde percentages van de antwoorden "weinig verbeterd", "niet verbeterd" en "verslechterd". Hieruit blijkt deze trend duidelijk. Na ont-

slag is steeds een toename van ca. 10% van de patiënten die zeggen dat de problemen weinig/niet verbeterd of verslechterd is. Bij de derde follow-up is dit verschil nog slechts enkele procenten.

Tabel 8.20 Resultaten op de vier meetmomenten van het eerste probleem. Som van de percentages van de drie ongunstigste antwoordcategorieën.

Meetmoment	1981	1982	1983	1984	1985	T
ontslag	26	29	36	19	39	28
1e follow-up	37	42	45	29	*	38
2e follow-up	34	46	47	32	*	39
3e follow-up	32	30	42	21	*	31

*) geen gegevens meer bekend.

De globale effectvraag: "hoe voel je je nu?" wordt eveneens op alle drie de follow-up momenten gesteld. De resultaten corresponderen in grote mate met de probleemgebonden evaluatie, zij het dat dit subjectievere oordeel iets beter is dan de beoordelingen van de verbetering van het geformuleerde probleem.

Het oordeel: ik voel met "niet zo goed" of "slecht" wordt bij de eerste follow-up door 36% gegeven. Dezelfde trend is hier aanwezig: 1983 is de negatieve uitschieter met 52%. In tabel 8.21 zijn weer de vier meetmomenten weergegeven, waarbij de twee meest negatieve categorieën samen zijn gepresenteerd.

Tabel 8.21 Resultaten op de vier meetmomenten op de vraag "hoe voel je je nu over het algemeen".
(% van de twee negatieve antwoorden:
"niet zo goed" en "slecht" samen.)

Meetmoment	1981	1982	1983	1984	1985	T	(N)
ontslag	24	19	39	13	27	23	(326)
1e follow-up	29	37	52	28	*	36	(266)
2e follow-up	30	37	37	27	*	32	(245)
3e follow-up	28	26	40	25	*	30	(210)

*) geen gegevens meer bekend

Ook nu blijkt een stijging van de twee negatieve antwoorden bij de eerste follow-up met gemiddeld 13%, die in het verdere verloop van de follow-up weer wat afneemt. Hier kan wederom de vraag gesteld worden of de beoordeling bij ontslag positief vertekend is of dat deze daling werkelijk zo groot is. Daarom is sinds ontslag bij de 1e follow-up een vraag opgenomen om achteraf aan te geven hoe men zich bij ontslag voelde. Wordt nu het antwoord bij ontslag en bij de eerste follow-up vergeleken, dan blijkt enerzijds dat de samenhang zeer groot is ($\chi^2 = 140.2$ $p < .001$), 63% geeft precies hetzelfde oordeel, en 29 % wijkt slechts één antwoordcategorie af. Degenen die een afwijkend oordeel geven, zeggen retrospectief vaker dat het bij ontslag slechter ging dan op het ontslagmoment zelf: het middenoordeel daalt van 44% naar 31%, terwijl de twee laagste antwoorden elk stijgen met respectievelijk 3% en 6%. Dit lijkt een sterke aanduiding voor een positieve antwoordtendentie bij ontslag. Een bewijs is dit evenwel niet omdat er retrospectief ook een vertekening van de herinnering kan optreden, met name door de gebeurtenissen in de tussenliggende periode i.c. een verslechtering van de status quo.

Er worden nog twee andere vragen gesteld over het oordeel achteraf over de behandeling op Overwaal: "Vind je achteraf dat je problemen bij ontslag voldoende verbeterd waren" en een algemeen satisfactie-oordeel over opname en verblijf in Overwaal. De antwoorden op de eerste vraag laten zien, dat 53% zegt dat de problemen voldoende verbeterd waren, 11% twijfelt hierover en 35% vindt de problemen te weinig verbeterd. De algemene satisfactie-vraag laat een lagere tevredenheid zien vergeleken met dezelfde vraag bij ontslag: het percentage zeer tevreden blijft gelijk met ca. 40%, terwijl het antwoord "tevreden" daalt van 48% naar 37%. Het percentage patiënten dat zegt het niet te weten of (zeer) ontevreden te zijn, stijgt van 11% naar 24%. Al deze vragen gezamenlijk lijken de conclusie te rechtvaardigen, dat er sprake is van een zekere positieve antwoordtendentie bij ontslag, hetzij bewust - hoewel de lijsten betrouwbaar waren -, hetzij onbewust omdat men niet wist hoe het na ontslag zou gaan, men was immers meerdere maanden klinisch behandeld en misschien te optimistisch over de periode erna.

De overige vragen gaan over de eventuele hulpverlening na ontslag de hulpbehoefte van de patiënt op dat moment. In tabel 8.22 staat een overzicht van de verkregen of gewenste hulpverlening op de drie follow-up momenten. Hieruit blijkt, dat een jaar na ontslag nog 57% hulp heeft (inclusief 16% die andere hulp wenst) en 13% hulpverlening zoekt. Een stijgend percentage, van 15% naar 30%, heeft geen hulp meer nodig.

Over de manier waarop door Overwaal de voortgezette hulp is geregeld, is 75% (zeer) tevreden. Na een intensieve klinische hulp is het wenselijk, dat zeer snel ambulante hulp start. Bij 37% gebeurt dit binnen een week na ontslag, bij 29% na 2 à 3 weken. Bij 25% duurt dit 1 tot 3 maanden, terwijl bij de overige 8% 3 maanden na ontslag nog geen hulp is gestart.

Tabel 8.22 Overzicht van de verkregen of gewenste hulp op de drie follow-up momenten (%). Vraag: Zou je op dit moment (andere) hulp willen hebben?

Antwoord:	1e	2e	3e
Nee, ik heb al goede hulp	50	50	41
Nee, ik heb geen hulp nodig	15	18	30
Ja, ik zou hulp willen hebben	17	12	13
Ja, ik zou andere hulp willen hebben	18	19	16
aantal patiënten	246	229	194

Bij de tweede follow-up zegt 18% dat er problemen zijn geweest met de hulpverlening na ontslag, met name dat het te lang duurde voordat deze startte. Op dat moment staan 3 personen nog op een wachtlijst, na een jaar is dit er nog 1.

De follow-up periode van 1 jaar overziend, blijkt een sterke terugval bij 3 maanden na ontslag met een lichte verbetering bij 6 en 12 maanden. Achteraf is de tevredenheid over Overwaal minder hoog dan bij ontslag. Na een jaar heeft of wenst nog 70% hulpverlening. Op dat moment zegt ruim 30% een uitkering te ontvangen (mede) op grond van hun problemen, en wordt nog 55% in wisselende mate tot zeer ernstig belemmerd in hun dagelijks functioneren. Hun algemene situatie beoordelend zegt dan 38% dat het goed gaat, 32% zegt "redelijk" en 30% "niet goed" of "slecht".

8.5.5 Follow-up vragenlijst hulpverlener.

In de follow-up vragenlijst die naar de voortgezette hulpverlener word gezonden, wordt gevraagd naar de manier waarop deze hulp is geregeld (1e follow-up), de actuele hulpverleningssituatie en beoordeling van het functioneren van de cliënt.

Om met dit laatste te beginnen: aan de hulpverlener werd dezelfde vraag voorgelegd als aan de cliënt: "Hoe gaat het nu", met dezelfde antwoordcategorieën, alsmede vragen over de vorderingen sinds ontslag en het niveau van functioneren. Conform tabel 8.21 worden in tabel 8.23 de antwoordpercentages van de twee onderste antwoordcategorieën samen vermeld op de vraag "Hoe gaat het nu" op de drie follow-up momenten door de hulpverlener beoordeeld. Worden de oordelen van behandelaar en cliënt met elkaar vergeleken, dan is de overeenstemming niet groot. De beoordelingen van de therapeut zijn over het algemeen positiever.

De vraag of er verbetering is opgetreden na ontslag c.q. sinds de start van de voortgezette hulpverlening laat wel een duidelijk beeld zien: bij de eerste follow-up zegt 26% te

Tabel 8.23 Resultaten op de drie follow-up momenten op de vraag "Hoe gaat het nu". Som van de percentages van de twee meest negatieve antwoordcategorieën "niet zo goed" en "slecht" samen.

Meetmoment	1981	1982	1983	1984	T	(N)
1e follow-up	26	30	35	18	27	(196)
2e follow-up	33	28	38	36	32	(174)
3e follow-up	11	12	25	28	19	(140)

vinden dat de patiënt (veel) beter is geworden, 14% zegt dat deze slechter is geworden. De overige 60% zegt "wel wat beter" of "hetzelfde" gebleven. Bij de derde follow-up is 43% (veel) beter geworden, slechts 4% slechter en 52% "wel wat beter" of "hetzelfde". De beoordeling van de therapeuten is ook hier aanmerkelijk positiever dan die van de patiënten, zoals uit tabel 8.24 blijkt. De trend is hier wel overeenkomstig en bij de 3e follow-up zijn de verschillen nog gering.

Tabel 8.24 Mate van verbetering na ontslag door patiënt (P) en therapeut (T) beoordeeld bij de drie follow-up momenten.

Verbetering	1e f.-up		2e f.-up		3e f.-up	
	P	T	P	T	P	T
(veel) beter	29	26	23	37	38	43
wel wat beter/hetzelfde	40	60	55	54	48	53
(veel) slechter	31	14	21	9	14	4

De vraag over het niveau van functioneren laat een grotere overeenstemming zien: bij de 3e follow-up zeggen beiden dat ca. 50% van de patiënten (bijna) normaal functioneert, 13% zegt dat de patiënt nauwelijks of niet kan functioneren. Bij de patiënten is dit 12%. De beoordelingen over de motivatie en de hulpverleningssituatie zijn onveranderlijk positief; ca. 70% zegt dat deze goed tot zeer goed is.

Bij de eerste follow-up wordt ook aan de hulpverlener gevraagd wat zijn oordeel is over het regelen van de nazorg en de zinvolheid van de opname op Overwaal. Zoals eerder vermeld, is er een afspraak dat voortgezette hulpverlening meer dan 2 maanden vóór ontslag geregeld zou moeten worden. Dit gebeurt slechts bij 5% van de patiënten. Wordt als cri-

terium gehanteerd, dat minimaal 3 weken voor ontslag deze hulp geregeld dient te zijn, zoals dit bij de ontslagvragenlijst van de psychotherapeut gevraagd wordt, dan voldoet 60% hieraan. Bij 35% is later of nooit (5%) hulp geregeld. Dit resultaat verschilt van het antwoord van de psychotherapeut, die zegt dat hij bij 80% deze hulp drie weken voor ontslag geregeld heeft.

In totaal is 75% wel tevreden over de manier waarop deze voortgezette hulp geregeld is, hetgeen overeenkomt met het oordeel van de patiënt. Ruim 5% zegt nooit informatie (uitschrijfbrief, e.d.) te hebben ontvangen van Overwaal. Over de inhoud van de verkregen informatie luiden de oordelen: uitgebreid/goed zegt 90%, (zeer) overeenstemmend met eigen indrukken en (zeer) nuttig antwoordt steeds 74%. Achteraf vindt 86% dat deze patiënt terecht verwezen is. Deze oordelen zijn over het geheel goed te noemen.

Het resultaat van de opname in Overwaal wordt door de hulpverlener bij de eerste follow-up wisselend beoordeeld. Uit tabel 8.25 blijkt dat het oordeel "goed" stijgt van 15% naar 25% in 1984. Met name het jaar 1982 laat een groot percentage "ik twijfel" en "geen oordeel" zien van samen 45%.

Tabel 8.25 Beoordeling resultaat opname in Overwaal door de voortgezette hulpverlener bij de 1e follow-up (%)

Antwoord	1981	1982	1983	1984	T
goed	15	16	16	25	18
redelijk	44	34	56	53	47
ik twijfel sterk	21	34	19	16	22
niet goed/slecht	14	5	7	6	8
geen oordeel	6	11	2	-	5
aantal patiënten	52	44	43	55	194

Worden de oordelen van de hulpverlener over de opname op Overwaal vergeleken met die van de patiënt bij de eerste follow-up, dan blijkt dat laatstgenoemde vaker een positief oordeel geeft; ca. 76% zegt (zeer) tevreden te zijn, bij de hulpverleners is dit 65%.

Samengevat zijn de oordelen van patiënt en hulpverlener betrekkelijk uniform, zij het dat de hulpverlener positiever is over het functioneren van de patiënt dan deze dat zelf is. Dezelfde trend is evenwel herkenbaar: een verminderende beoordeling van de status quo over de periode 1981-1983, met een verbetering in 1984. De beoordeling van de behandeling op Overwaal laat over 1983 en 1984 een stijging zien in vergelijking met 1981 en 1982.

In paragraaf 8.1 is gemeld dat de ontslag- en follow-up-vragenlijsten die gedurende de gehele evaluatieperiode zijn gebruikt, in 1980 op enkele onderdelen zijn veranderd. Bij de beschrijving van de resultaten is aanvankelijk getracht alle gegevens te transponeren tot één bestand en de evaluatieperiode in zijn geheel c.q. in jaargroepen (conform hoofdstuk 7) te presenteren. Echter, de veranderingen die in de lijsten zijn aangebracht, maakten dit onmogelijk. Hier zullen deze verschillen worden weergegeven, waarmee tevens wordt gedemonstreerd dat kleine tekstuele aanpassingen belangrijke gevolgen kunnen hebben.

In par. 8.3 en 8.4 zijn de resultaten van twee globale effectvragen weergegeven. Deze luiden: "Hoe voelt u zich nu?" en "Wat is het verschil tussen opname en ontslag?". Deze vragen zelf bleven hoegenaamd ongewijzigd in de lijsten gehandhaafd, alleen het aantal antwoordmogelijkheden werd met één uitgebreid. Hieronder staan de twee vragenvormen weergegeven. Bij de tweede vraag staan de wijzigingen tussen haakjes.

1976 - 1980	1981 - 1985
Hoe voelt u zich nu m.b.t. het probleem waarvoor u werd opgenomen.	Hoe voel je je nu over het algemeen?
1. Goed, problemen zijn over	1. Uitstekend
2. Redelijk, problemen zijn minder	2. Goed
3. Niet goed, problemen bestaan nog steeds	3. Gaat wel
4. Slecht, problemen zijn ernstig	4. Niet zo goed
	5. Slecht

Als u (je) zich (je) nu vergelijkt met het moment dat u (je) opgenomen werd op Overwaal, wat is dan uw (je) oordeel:

1. Ik voel me nu veel beter dan bij opname
2. Ik voel me nu beter dan bij opname)
2. (3.) Ik voel me (nu) wel wat beter dan bij opname
3. (4.) Ik voel me (nu) hetzelfde als bij opname
4. (5.) Ik voel me (nu) slechter dan bij opname.

De resultaten over de periode 1977-1985 staan in tabel 8.26 en 8.27 vermeld.

Uit tabel 8.26 blijkt, dat het toevoegen van één antwoord aan het positieve uiteinde van de schaal betekent, dat het grote percentage dat voorheen "redelijk" antwoordde, verdeeld wordt over het antwoord "goed" en "gaat wel". Het percentage antwoorden "niet zo goed" daalt, wellicht ook door de relatieve positieverandering van dit alternatief. Ook het wegla-

Tabel 8.26 Resultaten op de vraag: "Hoe gaat het nu?" (%).

antwoord:	77/78	79/80	81/82	83/84	85
uitstekend	-	-	3	4	1
goed	5	7	34	26	37
redelijk/gaat wel	74	63	42	38	30
niet zo goed	15	25	16	15	22
slecht	5	5	5	10	9
aantal patiënten	146	117	122	135	64

Tabel 8.27 Resultaten op de vraag naar het verschil tussen opname en ontslag over de totale periode (%).

antwoord:	77/78	78/80	81/82	83/84	85
veel beter	52	38	24	21	14
beter	-	-	24	36	37
wel wat beter	37	47	35	27	25
hetzelfde	5	5	11	4	8
slechter	6	9	7	12	9
aantal patiënten	147	117	121	135	64

ten van de specificaties zoals "problemen zijn minder" zal invloed hebben gehad. De verschuivingen lijken als volgt te zijn geweest:

- Goed -> uitstekend, enkelen goed;
- Redelijk -> 50% goed en 50% gaat wel;
- Niet zo goed -> 40% gaat wel en 60% niet zo goed;
- Slecht -> slecht.

Bij de vraag naar het verschil tussen opname en ontslag zijn de veranderingen geringer. Alle formuleringen zijn identiek, alleen is een nieuw alternatief tussengevoegd. De verschuivingen zijn nu ook minder groot: een gedeelte van het percentage van het antwoord "veel beter" verschuift nu naar "beter", evenals een gedeelte van "wel wat beter". Voor het overige lijkt het antwoordpatroon overeenkomstig.

Gelijke verschuivingen doen zich ook voor bij de follow-up vragenlijsten: bij de 1e follow-up antwoordde een sterk dalend percentage, van 67% in 1978 naar 31% in 1980, "redelijk", terwijl antwoordpercentage van "niet zo goed" steeg van 22% naar 48% in 1980. Bij toevoeging van het antwoord "uitstekend" in 1981, antwoordde in dat jaar 35% "goed", 35% "gaat wel" en 16% "niet zo goed".

Bij de vraag naar het behandelingsresultaat t.a.v. de geformuleerde problemen bij opname doet zich hetzelfde verschijnsel voor. Ook hier is een antwoordcategorie tussenge-

voegd; het antwoord "het probleem is tamelijk goed verdwenen" is vervangen door twee antwoorden: "voor het grootste gedeelte" en "min of meer verbeterd". De overige antwoorden zijn niet gewijzigd. Uit de resultaten blijkt dat bij de oude lijst "tamelijk goed" en "een beetje verdwenen" nu over drie antwoorden worden verdeeld (zie tabel 8.28). Ook nu blijven de overige antwoordpercentages constant.

Tabel 8.28 Resultaten op de vraag naar het behandelings-effect op de problemen over de totale periode (%)

antwoord:	77/78	79/80	81/82	83/84	85
geheel verdwenen	8	8	7	4	3
grotendeels	x	x	28	30	31
tamelijk goed	57	43	x	x	x
min of meer	x	x	36	39	27
een beetje verdwenen	29	36	17	15	22
hetzelfde gebleven	3	9	9	10	11
erger geworden	3	4	3	3	6
aantal patiënten	146	116	122	135	64

Op het criteriumprobleem bij de specifieke effectvraag (is het antwoord: "een beetje verbeterd" een voldoende mate van verbetering) wordt nu in een ander licht geplaatst. Het relatief hoge percentage patiënten dat een te gering effect zou melden, daalt bij de nieuwe vraagvorm sterk. Over de periode 1981-1984 meldt nog ca. 30% dat de problemen weinig of niet zijn verbeterd of slechter zijn geworden, om in 1985 weer te stijgen naar 40%. Dat hier inderdaad sprake is van een verschuiving op grond van een andere vraagstelling blijkt ook uit de resultaten op vragen die wel identiek zijn. Zo geeft de vraag over de tevredenheid over de opname op Overwaal bij de 1e follow-up gedurende alle jaren wel een stabiel antwoordpatroon, afgezien van de eerder beschreven trend van een lichte daling van het positieve oordeel van 83% in 1977 naar 70% in 1983 met weer een stijging naar 83% in 1984.

Hiermee is voldoende gedemonstreerd hoezeer het aantal antwoordmogelijkheden van invloed is op de antwoordpercentages, ook bij gelijk blijven van de omschrijvingen bij ieder antwoord. Er wordt tevens aangetoond, dat de wijzigingen juist waren: er bestaat nu een betere spreiding van de antwoordpercentages over de vijf antwoorden. Dit alles rechtvaardigt de presentatie van de resultaten in twee afzonderlijke perioden en maakt het moeilijk over de totale periode conclusies te trekken. Per periode zijn de resultaten wel interpreteerbaar, zeker t.a.v. de gesignaleerde trends. Ook kan over elke periode afzonderlijk gezocht worden naar ver-

banden tussen de resultaten en andere gegevens (zie hoofdstuk 9). In de volgende paragraaf zullen de resultaten worden samengevat in de vorm van evaluaties van doelstellingen, zoals dat programma-evaluatie betaamt.

8.7 EVALUATIE VAN DOELSTELLINGEN.

In de voorafgaande paragrafen zijn de resultaten hoofdzakelijk bivariaat gepresenteerd. Van de vragen die in de verschillende evaluatieperioden door de patiënten en hun hulpverleners zijn beantwoord, zijn de antwoordpercentages gepresenteerd per jaar en slechts summier van commentaar voorzien. De presentatie was voornamelijk de status quo beschrijvend.

Primair is programma evaluatie een evaluatie van doelstellingen. Deze doelstellingen kunnen, zoals in hoofdstuk 3 vermeld is, "inspannings"- of "effect"-variabelen zijn: input- of output resultaten.

De doelstelling van Overwaal is (cf. hoofdstuk 5): "... patiënten opnemen met psycho-sociale problemen, die niet of niet voldoende behandeld kunnen worden bij de ambulante voorzieningen voor hulpverlening. Het behandelingsdoel is het bereiken van zodanige veranderingen dat de patiënt zonder verdere hulp zijn leven kan inrichten ofwel met vrucht (weer) gebruik kan maken van de ambulante voorzieningen". Sinds 1983 is de passage: "zonder verdere hulp zijn leven kan inrichten" uit de doelstelling geschrapt.

De input-gegevens zijn beschreven in hoofdstuk 7. De opgenomen patiënten voldoen aan de gestelde doelgroep:

- ca. 60% van de patiënten behoort tot de drie laagste sociale klassen;
- de leeftijd ligt tussen 18 en 60 jaar,
- de problematiek wordt hoofdzakelijk als "neurotisch" gediagnostiseerd
- de beoogde behandelingsduur van minder dan 5 maanden wordt door 60% bereikt, inclusief 20% (te) kort opgenomenen (minder dan 2 maanden).

De outputgegevens zijn in dit hoofdstuk beschreven. Aan de formele behandelingsdoelstelling voldoet strikt genomen 77% van de patiënten; 16% kan zonder hulp weer zijn leven inrichten en 61% denkt het met nazorg wel weer aan te kunnen. In hoeverre dit een verbetering is in vergelijking met de situatie vóór opname, is niet duidelijk. Op dat moment had 2% geen hulpverlening en 10% slechts een half jaar of minder. In het algemeen gesproken was er voor opname reeds een bepaalde vorm van hulpverlening, welke na ontslag gecontinueerd werd. Wil het behandelingsresultaat goed geëvalueerd kunnen worden, dan dient ook de kwaliteit van de verbetering, welke door de behandeling zou moeten ontstaan, in deze doelstelling gespecificeerd te zijn. Deze kwaliteit kan geoperationaliseerd

worden in de eerdergenoemde globale en specifieke, doelgerichte effectvragen. De resultaten op deze effectvragen werden jaarlijks in het jaarverslag vermeld. Over het beoogde effectniveau werd geen uitsluitend gegeven. Bij het ontbreken van deze absolute norm kan naar een relatieve norm worden gezocht. Door het presenteren van de gegevens over 9 opeenvolgende jaren ligt het voor de hand de jaren onderling te vergelijken, zoals in de beschrijving van de resultaten is gebeurd.

De totale evaluatieperiode (1976-1985) overziend kan gesteld worden dat:

- een stijgend percentage een gering behandelingseffect aan geeft: ca. 20% in 1977 tot 39% in 1985 zegt dat het opnameprobleem weinig/niet verbeterd of zelfs verslechterd is. Het gaat "niet zo goed" of "slecht" wordt door 19% in 1977 geantwoord, in 1985 is dit 31%, ondanks de verruiming van de antwoordcategorieën;
- de resultaten bij follow-up geven een overeenkomstig beeld, waarbij de 1e follow-up het meest ongunstige effect laat zien: 40% zegt in 1977 dat het niet goed/slecht gaat, in 1983 is dit geleidelijk geklommen tot 52% om in 1984 weer te dalen tot 28%;
- de tevredenheid over de behandeling is constant zeer goed. Ca. 70-90% is (zeer) tevreden over diverse aspecten van de behandeling c.q. behandelaars;
- het regelen van nazorg is bij 80% van de patiënten drie weken vóór ontslag gebeurd, maar volgens de voortgezette hulpverleners is dit slechts bij 60% van de patiënten het geval. Laat geregelde hulp leidt tot een ongewenst hiaat in de curatieve zorg, slechts bij ca. 30% sluit de voortgezette hulp direct aan op de klinische periode;
- in de follow-up periode blijken de problemen direct na ontslag sterk te verslechteren, waarna deze bij follow-up na 6 en 12 maanden weer geleidelijk verbeteren;
- een jaar na ontslag heeft 30% geen hulp meer nodig. De overige 70% heeft wel hulp of zou deze wensen.

Doorredenerend zijn dit enkele evaluaties welke naar aanleiding van de doelstelling van Overwaal gegeven kunnen worden. Zoals gezegd is de doelstelling zelf te vaag en teveel gericht op de hulpverlening na ontslag. Het weer "met vrucht gebruik kunnen maken van ambulante hulp" is een doel, dat moeilijk operationaliseerbaar is. Als deze hulp om externe redenen zoals wachtlijst, ziekte hulpverlener, verhuizing patiënt, e.d. niet of niet optimaal wordt gerealiseerd, is formeel niet aan de doelstelling voldaan. Maar dit had Overwaal niet (altijd) kunnen voorkomen. De doelstellingen dienen daarom ook op het behandelingsresultaat zelf te zijn gericht. Het ontbreken van concrete doelen maakt, dat dit programma-evaluatie-project descriptiever is dan het zou moeten zijn. Deze en andere kanttekeningen welke bij de evaluatie van dit project en het gebruik van de resultaten ervan te maken zijn, komen expliciet aan bod in hoofdstuk 11, waarin het programma-evaluatie-project zelf wordt geëvalueerd.

Hoofdstuk 9:

De resultaten geanalyseerd

...De statistiek, die wanneer zij met kennis van zaken beoordeeld en aangewend wordt, een krachtig en onwaardeerbaar middel is tot het aan het licht brengen van veel gewichtige waarheden, welke zonder haar verholen zouden blijven.
Schröder van der Kolk en Feith (1984).

9.1 INLEIDING.

In hoofdstuk 8 is voornamelijk descriptief beschreven hoe gedurende de periode 1974-1985 diverse aspecten van het functioneren beoordeeld zijn door de patiënten en hun hulpverleners. Deze aspecten betreffen vooral het resultaat van de behandeling, de tevredenheid over de opname en hiermede verband houdende zaken, alsmede de regeling van nazorg. Tal van gegevens zijn bivariaat weergegeven, het resultaat is gerelateerd aan het ontslagjaar. In incidentele gevallen zijn vergelijkingen gemaakt tussen gegevens onderling zoals die tussen therapeut en patiënt op hetzelfde meetmoment. Verdere verbanden of interrelaties zijn niet weergegeven.

Hiermee is aan het eerste doel van programma-evaluatie voldaan: deze informatie (periodiek) te presenteren, zodat deze als feedback kan dienen voor het beleid en dit zonodig kan worden veranderd indien de resultaten hiertoe aanleiding geven. Zoals eerder gesteld (hoofdstuk 3), is het wenselijk naast een globale evaluatie ook zo mogelijk een gedifferentieerde feedback te geven: in hoeverre leiden verschillen tussen patiëntenvariabelen, behandelingsvormen, e.d. tot verschillen in resultaten. Informatie uit dergelijke deels exploratieve deels toetsende analyses kan behulpzaam zijn bij het zoeken naar verklaringen voor de verkregen resultaten. Deze analyses zijn echter alleen relevant als de variabelen aan psychometrische kwaliteiten voldoen. Dergelijke psychometrische kenmerken hebben o.a. betrekking op de betrouwbaarheid en validiteit van de variabelen. Zo werd in hoofdstuk 8 herhaaldelijk melding gemaakt van mogelijke antwoordtendenties, met name bij de ontslagvragenlijst. Of deze daadwerkelijk bestaan, zou onderzocht kunnen worden. Het kruisen van dezelfde vragen van twee informanten - patiënt en hulpverlener - op hetzelfde tijdstip kan indicaties over de betrouwbaarheid en validiteit opleveren. In paragraaf 9.2 worden enkele van deze relaties tussen vragen beschreven.

Met behulp van cluster- en/of factoranalyses kan een zeer uitgebreid gegevensbestand worden gecomprimeerd. Enerzijds levert dit informatie op over de verbanden tussen de vragen, anderzijds kunnen de geëxtraheerde factoren gebruikt worden

om onderzoek te doen naar de oorzaken van resultaten. Homogeen samengestelde factoren, bestaande uit meerdere variabelen zullen betrouwbaardere informatie bevatten en daarom meer bruikbaar zijn om relaties met andere gegevens te leggen en trends in de resultaten op te sporen en op significantie te toetsen.

In paragraaf 9.3 wordt deze comprimering van de variabelen in factoren beschreven. Deze factoren worden vervolgens in paragraaf 9.4 gebruikt om middels multivariate variantie-analyses verbanden tussen de resultaten en andere gegevens te toetsen. Zoals in hoofdstuk 8 zijn de twee bestanden apart geanalyseerd: 1977-1980 en 1981-1984. De beschrijving concentreert zich evenwel op het tweede bestand, omdat dit meer uniform is opgezet: patiënt en therapeut beantwoorden dezelfde vragen op drie of vier opeenvolgende momenten. Ook de redactie van de vragen is evenwichtiger. Er zal regelmatig vermeld worden of de gevonden verbanden afwijken van die uit de periode 1977-1980.

9.2 BETROUWBAARHEID.

In de testtheorie worden verschillende vormen van betrouwbaarheid onderscheiden: interne betrouwbaarheid ofwel homogeniteit en externe betrouwbaarheid, relatie tussen twee of meer metingen, door dezelfde persoon of door verschillende personen (test-hertest en interbeoordelaarsbetrouwbaarheid) (Drenth, 1972).

9.2.1 Interne consistentie.

De interne consistentie geeft aan in hoeverre de items dezelfde informatie geven. Om deze itemhomogeniteit na te gaan, is eerst noodzakelijk te onderzoeken of de vragen uit één domein komen (De Bruyn, 1985, p. 175). De hier ontwikkelde vragenlijsten zijn niet als zodanig bedoeld, gepoogd is zo gevariëerd mogelijke informatie te verwerven over tal van aspecten van het functioneren. Niettemin kunnen groepen van vragen bij nadere analyse een grote onderlinge samenhang hebben, zodat er van subschalen gesproken kan worden. Per subtest kan vervolgens de homogeniteit onderzocht worden. Het is ook niet uitgesloten, dat de somscores op deze subschalen onderling grote samenhang vertonen, doordat de domeinen met elkaar gerelateerd zijn. Dit kan bijv. het geval zijn met subschalen als satisfactie en effect.

Per vragenlijst in zijn geheel zijn de betrouwbaarheid-scoëfficiënten (Cronbach's α) berekend alsook per subschaal, zoals deze op basis van de inhoud van de vragen zijn samen te stellen. In tabel 9.1 staan deze coëfficiënten (α) vermeld. Deze zijn berekend over de vragen die dimensionele informatie bevatten; de nominale variabelen zijn dus uitgesloten. Uit deze tabel blijkt, dat alle vragenlijsten zeer hoge

betrouwbaarheidscoëfficiënten hebben (α is ca. 0.85). In zoverre is er kennelijk sprake van homogene en consistente informatie. Ook van de subschalen zijn de coëfficiënten hoog, bij testverlenging tot de lengte van de hele lijst zouden deze waarden bereiken van 0.95. (Lord en Novick, 1968).

Tabel 9.1 Betrouwbaarheidscoëfficiënten (α) van de vragenlijsten en de subschalen.

Lijst	n	N*)	α
ontslaglijst patiënt	33	185	0.86
- satisfactie	8	243	0.87
- psychotherapeut	4	243	0.81
- sociotherapeut	5	243	0.83
- effect	4	243	0.81
Ontslaglijst therapeut	28	286	0.76
Follow-up patiënt 1	13	149	0.88
2	8	164	0.86
3	8	124	0.89
Follow-up nazorg 1	14	123	0.81
2	5	172	0.86
3	5	129	0.85

n = aantal vragen, N = aantal personen, α = KR20

*) het aantal personen wordt bepaald door degenen die van een lijst alle vragen hebben beantwoord.

Bij de therapeut-ontslagvragenlijst hebben enkele beoordelingen van de intakecriteria een zeer kleine variantie en een lage of negatieve correlatie met de totaalscore. Deze criteria zijn bij verdere analyse uitgesloten. Dit zijn: leeftijd, intelligentie, organiciteit, somatische ziekte, psychotisme, narcomanie, suicidaliteit, aanpassing in leefmilieu, coöperatie eigen milieu, terugplaatsbaarheid en noodzaak klinische hulp.

Geconcludeerd kan worden, dat de interne consistenties van de vragenlijsten goed zijn en de resultaten in dat opzicht betrouwbaar genoemd kunnen worden. Opsplitsing in subschalen op basis van de inhoud van de vragen doet de betrouwbaarheid per subschaal stijgen, als gecorrigeerd wordt voor testlengte.

9.2.2 Interbeoordelaarsbetrouwbaarheid.

Interbeoordelaarsbetrouwbaarheden zijn te berekenen door kruistabellering van dezelfde vragen die door patiënt en therapeut afzonderlijk zijn ingevuld, of door de patiënt op verschillende momenten. Hieronder volgt een opsomming van dergelijke vragen met de toetsing van het verband tussen de antwoorden.

- Tevredenheid over de behandeling op Overwaal, door de patiënt bij ontslag en eerste follow-up beoordeeld: $\chi^2 = 141.8$, $p < .001$, $N = 269$.
- Hoe veel(de) je je (bij ontslag), door de patiënt bij ontslag en eerste follow-up beoordeeld: $\chi^2 = 140.2$, $p < .001$, $N = 270$.
- Oordeel over de behandelingsduur, door de patiënt bij ontslag en eerste follow-up gegeven: $\chi^2 = 103.6$, $p < .001$, $N = 276$.
- Oordeel over de behandeling op Overwaal, door de patiënt en de hulpverlener bij de eerste follow-up gegeven: $\chi^2 = 34.4$, $p = 0.0001$, $N = 216$.
- Oordeel over de status quo bij de eerste follow-up volgens de patiënt en zijn hulpverlener: $\chi^2 = 67.4$, $p < .001$, $N = 220$.
- Idem bij tweede follow-up: $\chi^2 = 41.4$, $p < .001$, $N = 169$.
- Idem bij derde follow-up: $\chi^2 = 34.1$, $p < .001$, $N = 114$.
- De vragen over de verandering sinds ontslag door de patiënt en zijn hulpverlener beoordeeld geven dezelfde significante overeenstemmingen te zien.
- De vragen over het functioneren van de patiënt op de drie follow-up momenten door hemzelf en zijn hulpverlener beoordeeld, geven resp. χ^2 : 43.0, 20.2 en 20.0. De p-waarden zijn resp.: $p < .001$, $p = .02$ en $p = .02$. Hier wordt de overeenstemming geringer, maar blijft significant bij een grens van $\alpha = .05$. Deze geringere overeenstemming wordt veroorzaakt door het feit dat de hulpverleners gemiddeld het functioneren van de patiënt positiever beoordelen dan deze dat zelf doet.
- Minder overeenstemming wordt aangetroffen tussen de oordelen van de psychotherapeut van Overwaal over het regelen van de nazorg en de informatie van deze voortgezette hulpverlener. De psychotherapeut antwoordt, dat deze bij 84% van de patiënten 3 weken voor ontslag contact met de voortgezette hulpverlener heeft opgenomen, terwijl deze hulpverlener zegt dat dit bij 33% van de patiënten niet gebeurd is. Ook de toezegging bij 83% van de patiënten binnen 2 weken na ontslag de uitschrijfbrief te versturen blijkt - volgens de nazorger - niet altijd gerealiseerd te worden: 10% ontvangt deze nooit en 25% later. De betrouwbaarheid van de vragenlijsten c.q. de informanten is op basis van deze gegevens moeilijker te verifiëren, omdat een gedeelte van deze verschillen veroorzaakt kan worden door communicatieproblemen: verzonden informatie komt misschien niet aan bij de behandelaar die de follow-up lijst invult, of de

behandelaar krijgt de patiënt via een intake team en is niet op de hoogte wanneer met het bureau (RIAGG) contact is opgenomen.

Als laatste wordt de relatie nagegaan tussen de prognose bij ontslag (hoe zal het in de toekomst met de patiënt gaan), gegeven door de psychotherapeut en de beoordeling van de patiënt van de status quo bij ontslag en op de drie follow-up momenten. Deze relatie is eerder een aanduiding van de (predictieve) validiteit van deze prognose dan een maat voor de betrouwbaarheid. Deze verwachting van de psychotherapeut correspondeert significant met het oordeel van de patiënt over "hoe het gaat". Naarmate de tijd verstrijkt, daalt de relatie licht, maar het verband blijft significant aanwezig: de prognose van de psychotherapeut bij ontslag stemt ook bij de 3e follow-up nog met het oordeel van de patiënt overeen ($\chi^2 = 19.74$ df = 8 p = .01).

Samenvattend blijkt, dat de betrouwbaarheden van de gebruikte vragenlijsten goed genoemd kunnen worden.

9.3 REDUCTIE IN FACTOREN.

Het gehele bestand met de gegevens van de evaluatievragenlijsten bevat per patiënt 13 "records" met in totaal ca. 700 variabelen. Een dergelijk bestand is onhandelbaar en maakt comprimering noodzakelijk. Een geschikte procedure voor een dergelijke datareductie is factor-analyse.

Over de meest geschikte analysemethode verschillen de meningen. Harman (1967) adviseert hiervoor de principale assen analyse, terwijl Uberla (1971) aan de principale componenten analyse de voorkeur geeft. Deze laatste methode is ook hier gebruikt. Per vragenlijst is een principale componentenanalyse uitgevoerd met paarsgewijze weglating van ontbrekende waarnemingen, waarna een varimax rotatie werd verricht. Deze analyse is op alle niet categoriale (nominale) variabelen uitgevoerd en zo nodig zijn antwoorden gehercodeerd om variabelen unidimensioneel te maken. Berekeningen zijn verricht met behulp van SPSS-X (Norusis, 1986); tenzij anders vermeld, zijn de "default"-waarden gehanteerd.

9.3.1 Ontslagvragenlijst patiënt.

Er zijn 10 factoren geëxtraheerd, die gezamenlijk 69% van de variantie verklaren. De eerste factor heeft een eigenwaarde van 10,6 met 28,7% van de variantie. Op deze factor laden alle variabelen die betrekking hebben op het behandelingsresultaat, de tevredenheid en de beoordeling van de psychotherapeut. De overige negen factoren hebben eigenwaarden die variëren van 3.2 tot 1.01 met een percentage verklaarde variantie van 8.6 tot 2.7. Alle communaliteiten liggen tus-

sen .52 en .86. Na rotatie ontstaan de volgende factoren met hun variabelen en ladingen hoger dan .40. Tussen haakjes staan de factorladingen op de geroteerde factor vermeld.

1. Effect.

- Hoe voel je je nu (.81).
- Verschil tussen opname en ontslag (.80).
- Kun je beter met je problemen omgaan (.79).
- Beoordeling verbetering problemen (.75).
- Reden ontslag (.73).
- Ontslagtijdstip (.55).

2. Sociotherapeut.

- Tevredenheid (.89).
- Sloot aanpak aan (.85).
- Had je een andere willen hebben (.85).
- Bereikbaarheid (.66).
- Begeleiding leefmilieu (.42).

3. Psychotherapeut.

- Zou je 'n andere gehad willen hebben (.82).
- Tevredenheid over psychotherapeut (.81).
- Sloot aanpak aan (.76).
- Tevredenheid over kwaliteit behandeling (.56).
- Tevredenheid over soort behandeling (.45).
- Bevatte programma alles wat nodig was (.42).

4. Contact met medepatiënten.

- Intensiteit (.80).
- Frequentie (.76).
- Aangenaamheid (.75).
- Voelde je je thuis (.72).
- Hulp ontvangen (.64).

5. Algemene tevredenheid.

- Weer terugkomen (.77).
- Een ander adviseren (.67).
- In 't algemeen tevreden (.50).

De overige factoren bevatten minder dan drie variabelen en worden hier verder niet meer beschreven. Dubbelladingen komen zeer zelden voor. Opmerkelijk is, dat de frequentie van het contact met de psychotherapeut en de sociotherapeut samen op een factor laden, en dat deze niet gerelateerd zijn aan de tevredenheid over deze behandelaars. Bovenstaande 5 factoren worden in de multivariate analyses in paragraaf 9.4 gebruikt, waarbij de gemiddelde verbetering van de bij opname genoemde problemen ('target complaints') ook als aparte variabele is gebruikt.

9.3.2 Ontslagvragenlijst therapeut.

Als invoer voor de factor-analyse zijn naast de beoordelingen over de behandeling van de patiënt ook de beoordelingen van de intakecriteria gebruikt. Dit leidt evenwel tot een onjuiste correlatiematrix. Aangezien diverse intake-criteria een zeer kleine spreiding hebben, zijn deze van verdere ana-

lyse uitgesloten (zie ook paragraaf 9.2.1). Bij de resterende 18 vragen zijn 3 factoren geëxtraheerd die gezamenlijk 60,9% van de variantie verklaren. De eerste factor is in feite een algemene factor waar alle vragen op laden en verklaart 42% van de variantie. De overige twee factoren nemen elk resp. 11.2 en 7.8% van de variantie voor hun rekening. De communaliteiten variëren van .32 to .79. Door rotatie ontstaan de volgende factoren met de ladingen per variabele ($r > .40$).

1. Effect.

- Effect van de behandeling (.76).
- Intake criterium van behandelingsduur (.71).
- Verwachting na ontslag (.71).
- Intakecriterium persoonlijkheidsstoornis (.69).
- Motivatie (.66).
- Problemen voldoende opgelost (.65).
- Zinvolheid opname (.62).
- Intake criterium motivatie (.61).
- Intake criterium uitageren (.49).

2. Beëindiging behandeling.

- Behandeling afgebroken (.82).
- Tijdstip ontslag (.82).
- Behandeling juiste wijze afgerond (.80).

3. Contact met patiënt.

- Betrokken met problematiek (.82).
- Intensiteit contact (.72).
- Aangenaamheid contact (.71).
- Interesse voor problematiek (.68).
- Tevreden over relatie (.60).

Ten behoeve van verdere analyses zijn de intakecriteria welke op de eerste factor laden, apart genomen, waardoor 4 samengestelde variabelen ontstaan.

9.3.3 Follow-up patiënt.

De drie follow-up vragenlijsten zijn zowel gezamenlijk als elk apart geanalyseerd. Dit leidde tot hoegenaamd dezelfde factorstructuur. De vragen welke alleen bij de eerste follow-up gesteld zijn over de behandeling op Overwaal, vormen een aparte factor. In tabel 9.2 worden de drie factor-analyses per lijst gepresenteerd. Deze factoranalyses verklaren elk resp. 60%, 61% en 63% van de variantie. De communaliteiten liggen hoofdzakelijk tussen .50 en .80. Opmerkelijk is, dat de gemiddelde verbetering van de problemen bij de eerste follow-up zowel laadt op de factor "effect" als op de factor "oordeel Overwaal" (lading = .50). De vraag "hoe snel had je een afspraak met de voortgezette hulpverlener na ontslag" vormde een eigen factor samen met deze gemiddelde verbeteringsscore. Dit duidt op een relatie hiertussen. Bij verdere analyses zullen deze gemiddelde verbeteringsscores ook hier als aparte variabele worden gebruikt, naast de globale effect-score. De factor 2, met 2 variabelen, zal verder buiten beschouwing blijven.

Tabel 9.2 Factorstructuren van de drie follow-up lijsten welke de patiënt na 3, 6 en 12 maanden invult.

Factor en variabelen	Ladingen lijst		
1. Effect	1e	2e	3e
- hoe voel je je nu	.81	.86	.76
- verschil ontslag - nu	.79 *)	.86	.83
- verschil sinds vorige lijst	.91	.81	.82
- oordeel omgeving	-	.64	.81
- niveau van functioneren	.67	.58	.58
- gemiddelde verbetering problemen	.60	.94	.80
2. Hulp nu			
- tevredenheid	.91	.85	.82
- hulpbehoefte	.49	.73	.80
3. Oordeel Overwaal			
- status quo bij ontslag	.84		
- problemen voldoende verbeterd	.78		
- tevredenheid Overwaal	.66		
- duur opname	.61		
- oordeel omgeving	.57		

*) sinds opname

9.3.4 Follow-up hulpverleners.

De drie vragenlijsten welke de voortgezette hulpverlener invult, vertonen eveneens overeenkomstige factorstructuren. Bij iedere lijst blijkt de belangrijkste factor de beoordeling van de status quo te zijn. De eerste follow-up lijst bevat ook vragen over de regeling van de voortgezette hulpverlening. Deze blijken extra factoren te vormen. In tabel 9.3 staan de factoren en de factorladingen vermeld. De verklaarde varianties per lijst bedragen resp. 61.1%, 57.2% en 63.0%. Alle communaliteiten variëren van .45 tot .85. Opmerkelijk is, dat de beoordeling van de behandeling op Overwaal niet gerelateerd is aan de status quo, maar samen met de beoordelingen van de informatie van Overwaal een factor vormt.

9.4 SAMENVATTING.

De uitgevoerde factor-analyses laten elk duidelijk patroon zien, met (zeer) hoge ladingen op de factoren. De verklaarde variantie per factoranalyse zijn relatief hoog: 60% - 70%. Iedere lijst bevat een effect-factor met de vragen die betrekking hebben op het behandelingsresultaat, beoordelingen van de status quo, e.d. Daarnaast worden ook factoren gevonden met betrekking tot specifieke onderdelen van het behande-

Tabel 9.3 Factorstructuren van de follow-up lijsten
welke de hulpverleners na 3, 6 en 12 mnd. invult.

Factoren en variabelen	ladingen per lijst.		
	1e	2e	3e
1. Effect			
- hoe gaat het nu	.80	.84	.85
- verschil ontslag - nu	.80	.79	.78
- motivatie	.72	.71	.74
- tevredenheid hulpverlening	.77	.76	.80
- niveau van functioneren	.71	.67	.79
2. Regelen nazorg			
- juiste verwijzing	.66		
- wanneer geregeld	.66		
- tevredenheid over regeling	.65		
3. Informatie ontvangst			
- wanneer ontvangen	.84		
- oordeel tijdstip	.82		
4. Inhoud informatie			
- uitgebreidheid	.69		
- nuttigheid	.65		
- resultaat behandeling	.53		

lingsproces, zoals regelen nazorg, beëindiging behandeling, actuele hulpbehoefte. Bij de patiënt blijken aspecten als satisfactie, beoordeling psycho- en sociotherapeut, e.d. ook aparte factoren op te leveren. Deze factoren worden vervolgens gebruikt als nieuwe variabelen bij verdere analyses. Hiertoe wordt per factor een gemiddelde score berekend door de waarden van de variabelen waaruit de factor is gedefinieerd te sommeren en te delen door het aantal. Dit gemiddelde is alleen berekend wanneer de helft op meer van de variabelen een niet gemiste waarde hebben. Tenslotte wordt kort melding gemaakt van de factor-analyses van de vragenlijsten welke van 1976-1980 gebruikt zijn. De gevonden factoren vertonen grote overeenkomst met de hiervoor gepresenteerde. Bij de ontslagvragenlijst van de patiënt blijken eveneens de factoren: effect, beoordeling psychotherapeut en sociotherapeut, contact medepatiënten en beëindiging behandeling geëxtraheerd te worden. Ook de follow-up lijsten vertonen dezelfde factorstructuren. Deze resultaten onderbouwen de validiteit van de factoren en rechtvaardigen des te meer het gebruik ervan in de hierna volgende variantie analyses.

9.5 ANALYSE VAN EFFECTEN.

Na de comprimering van de vele variabelen middels factoranalyses zijn de resultaten gereduceerd tot 23 scores volgens bovenstaande procedure. Deze scores staan vermeld in tabel 9.4.

Tabel 9.4 Schalen, welke met behulp van de factor analyses zijn geconstrueerd.

Meetmoment	Door	Inhoud	Items
ontslag	patiënt	globaal effect	5
		verbetering doelen	1 - 5
		satisfactie	3
		sociotherapeut	6
		psychotherapeut	6
ontslag	psychotherp.	contact medepatiënt	5
		effect	5
		einde behandeling	3
follow-up	patiënt	globaal effect 1e f.u.	4
		globaal effect 2e f.u.	5
		globaal effect 3e f.u.	5
		verbetering doelen 1e f.u.	1 - 5
		verbetering doelen 2e f.u.	1 - 5
		verbetering doelen 3e f.u.	1 - 5
		beoordeling Overwaal	5
		hulpverlening 1e f.u.	3
		globaal effect 1e f.u.	5
		globaal effect 2e f.u.	5
follow-up	hulpverlener	globaal effect 3e f.u.	5
		regelen nazorg	3
		oordeel informatie	3
		ontvangst informatie	2

Zoals in hoofdstuk 2 is betoogd, dienen deze evaluatiere-sultaten eerder absoluut dan relatief geëvalueerd te wor-den. De berekende scores kunnen echter ook volgens de klas-sieke statistische analyse-technieken geanalyseerd worden. Het gaat dan om de beantwoording van de vraag of er bepaalde verbanden bestaan tussen bijv. de dimensies effect e.d. ener-zijds als afhankelijke variabelen en bepaalde onafhankelijke parameters anderzijds. Criterium georiënteerde evaluatie komt in de volgende paragraaf aan de orde.

De keuze van onafhankelijke variabelen is betrekkelijk willekeurig, wanneer er vooral explorerend te werk moet wor-den gegaan. Enkele voor de hand liggende relaties worden aan-gereikt door het onderzoek van Van der Veen en Brinkman

(1983). In dit artikel worden 20 min of meer voor de hand liggende moeilijkheidsvariabelen opgesomd, zoals ernst van de klacht, sociaal-economische situatie van de patiënt, e.d.

In het hier beschreven onderzoek zijn per therapeut voldoende patiënten aanwezig om ook de relatie tussen effect e.a. en de therapeutvariabele te toetsen. Daarnaast worden aard van de problematiek (diagnose), duur van de klachten en sociaal-economische gegevens als onafhankelijke variabelen gebruikt. In hoofdstuk 8 werd herhaaldelijk een trend gesignaleerd in de opeenvolgende evaluatie resultaten. Deze trend zal eveneens op zijn significantie onderzocht worden. Per onafhankelijke variabele worden de belangrijkste resultaten gepresenteerd. Deze betreffen de multivariate analyse met als toetsingsgrootte Hotellings T^2 en de univariate analyses met F-ratio's (Winer, 1962).

9.5.1 Psychotherapeut.

Op Overwaal werkten gedurende de hele evaluatieperiode 5 psychotherapeuten met voldoende patiënten om een analyse van de therapeutvariabele mogelijk te maken. Daarnaast waren steeds twee arts-assistenten werkzaam. Deze zijn als groep gezamenlijk genomen en worden in de verdere analyses als één therapeut beschouwd. Van alle 23 scores is nagegaan of er sprake is van een significant verschil tussen deze 6 psychotherapeuten. Therapeuten met zeer weinig patiënten zijn derhalve niet in de analyses betrokken. Bij de ontslagvragenlijst konden per therapeut 25-78 patiënten in de analyse worden betrokken. Per lijst volgen nu de resultaten van deze variantie-analyses.

Ontslagvragenlijst patiënt.

Multivariaat: alle scores gezamenlijk blijken een significant verband te hebben met de therapeut-variabele ($T^2 = .29$ $p < .001$). De univariate toetsen staan in tabel 9.5.

Tabel 9.5 Univariate variantie-analyse van de factoren van de patiënt ontslagvragenlijst met behandelend psychotherapeut als onafhankelijke variabele.

Factor	SS	MS	F	p
satisfactie	4.43	0.53	1.66	.144
gem. problemen	17.44	1.09	3.19	.008
sociotherapeut	5.70	0.42	2.74	.019
psychotherapeut	21.33	0.52	8.17	.000
globaal effect	2.75	0.47	1.16	.330
contact medepat.	3.68	0.57	1.30	.265

verklaring: SS = Sum of Squares; MS = Mean Square;
F = F-ratio; p = significantie.

Hieruit blijkt, dat er een significant verband (bij $\alpha = .01$) bestaat tussen de gemiddelde verbetering van de problemen en de beoordeling van deze psychotherapeut met de therapeut-variabele. Nadere beschouwing laat zien, dat de gemiddelde verbeteringsscore per therapeut varieert: drie therapeuten hebben elk een gemiddelde van 2.7, drie anderen (inclusief de arts-assistenten) echter van 3.2 (hoe lager hoe beter, schaal van 1 tot 6). Bij de beoordelingen van de tevredenheid over de therapeut variëren deze gemiddelden van 1,4 tot 2,1 met dezelfde volgorde als bij de effectmeting.

Analyse van de gegevens over 1977-1980 leveren dezelfde resultaten op. Ook dan blijkt er een significant verschil tussen dezelfde psychotherapeuten onderling op te treden.

Ontslagvragenlijst psychotherapeut.

De multivariate analyse resulteert ook hier in een significant resultaat: Hotellings $T^2 = 0.21$ $p < .001$. Bij de univariate analyse leiden de factoren: intakecriteria en contact met de patiënt tot significante resultaten (zie tabel 9.6).

Tabel 9.6 Univariate variantie-analyse van de factoren van de therapeut ontslagvragenlijst met de psychotherapeut als onafhankelijke variabele.

Factor	SS	MS	F	p
intakecriteria	15.56	3.11	5.48	<.001
effect	8.33	1.67	2.14	.065
beëindiging behandeling	12.92	2.58	2.61	.025
contact met patiënt	9.74	1.94	4.45	.001

voor verklaring van de afkortingen zie tabel 9.5

Interpretatie van dit resultaat is niet eenvoudig. Er blijkt een systematisch verschil te bestaan tussen de therapeuten ten aanzien van de intakecriteria en de beoordeling van de relatie met de patiënt. De patiënten worden over het algemeen naargelang plaatsruimte aan therapeuten toegewezen, zodat hier geen systematisch effect kan optreden. Kennelijk beoordelen de therapeuten zelf hun patiënten systematisch anders en bestaat er een belangrijke persoonlijke interpretatie van de eigen patiënten. Tevens blijkt, dat therapeuten geen verschil in effect aangeven, in tegenstelling tot hun patiënten, hetgeen niet zo verwonderlijk is. Iedere therapeut zal vanuit zijn eigen referentiekader het behandelings-effect beoordelen.

Follow-up vragenlijst patiënt.

Alle follow-up vragenlijsten zijn gezamenlijk en afzonderlijk geanalyseerd. Het verschil is vooral het aantal patiënten: bij analyse van de lijsten gezamenlijk worden alleen die patiënten betrokken die deze alle drie hebben ingevuld. Dit zijn 168 patiënten, terwijl bij de follow-up lijsten afzonderlijk van respectievelijk 228, 253 en 205 patiënten gegevens aanwezig zijn.

Alleen de gezamenlijke analyse geeft een significant resultaat te zien: de beoordeling van de patiënt over Overwaal bij het eerste follow-up moment ($SS = 7.35$ $MS = 1.47$ $F = 2.21$ $p = .05$). Deze beoordeling betreft de tevredenheid over het resultaat van de behandeling achteraf. Deze uitslag correpondeert met de significante verschillen bij de ontslag vragenlijst (zie tabel 9.5).

Follow-up vragenlijsten hulpverlener.

Analyses van de factoren van de follow-up vragenlijsten die de voortgezette hulpverlener heeft ingevuld, leveren geen enkele significante verschillen op tussen de psychotherapeuten van Overwaal.

Concluderend kan worden gesteld, dat de onafhankelijke variabele psychotherapeut tot belangrijke verschillen in effect aanleiding geeft. Dit is het duidelijkst aantoonbaar bij de beoordeling van de patiënt bij ontslag. Bij diverse factoren zijn de therapeuten consequent in te delen in meer of minder effectief, naar oordeel van de patiënt. Gedurende de follow-up periode zijn deze verschillen niet meer aanwezig.

9.5.2 Diagnose.

Gedurende de eerste weken van de opname stelt de psychotherapeut een of meer diagnoses volgens de I.D.C.-9. Als onafhankelijke variabele is de eerste diagnose genomen. Deze is doorgaans ook de belangrijkste. De veelheid van diagnoses (zie tabel 7.12) is in 6 groepen gecategoriseerd, met tussen haakjes het aantal patiënten met een dergelijke diagnose die ook de ontslagvragenlijst hebben ingevuld: angst (29), fobie (63), dwang (58), depressie (27), persoonlijkheid (13) en de overige (84). Per lijst volgen de resultaten van de variantie-analyses.

Ontslagvragenlijst patiënt.

Deze vragenlijst leverde geen significante resultaten.

Ontslagvragenlijst psychotherapeut.

Zowel multivariaat als univariaat blijken er significante verschillen te bestaan tussen de 6 diagnosegroepen: Hotelings $T^2 = 0.15$, $F = 2.35$, $p = .001$. De univariate verbanden staan in tabel 9.7. Hieruit blijkt, dat de factoren: beoordeling van de opnamecriteria, het effect en het contact met de patiënt significante verschillen geven tussen de diagnoses.

Tabel 9.7 Univariate variantie-analyse van de factoren van de therapeut ontslagvragenlijst met de diagnose als onafhankelijke variabele

Factor	SS	MS	F	P
opnamecriteria	16.59	3.32	5.96	< .001
effect	12.45	2.49	3.34	.006
beëindiging	6.56	1.31	1.31	.260
contact	8.44	1.69	3.90	.002

voor verklaring van de afkortingen zie tabel 9.5.

Het grootste verschil bestaat steeds tussen de diagnose persoonlijkheidsstoornis en de overige diagnoses. Hier kan sprake zijn van een contaminatie: wanneer een psychotherapeut een diagnose "persoonlijkheidsstoornis" stelt, spreekt hij tevens uit dat deze patiënt niet of zeer moeilijk behandelbaar is. Dit kan vervolgens in de beoordeling van het effect tot uiting komen. Patiënten die niet van deze diagnose op de hoogte zijn, geven immers geen verschil in behandelingsresultaat bij ontslag.

Follow-up vragenlijst patiënt.

Alleen bij de eerste follow-up vragenlijst van de patiënt bestaan significante verschillen tussen de diagnosegroepen. Multivariate analyse resulteert in Hotellings $T^2 = .18$, $F = 2.05$ $P = .004$. Univariate analyse leidt tot significante verschillen bij de factor "beoordeling Overwaal achteraf" en "actuele hulsituatie". Ook hier zijn de verschillen het grootst tussen de persoonlijkheidsstoornis en de overige diagnoses. De diagnose depressie geeft de meest positieve beoordelingen.

Follow-up vragenlijst hulpverlener.

In deze lijsten worden geen significante verschillen aangetroffen tussen de diagnoses.

Concluderend kan gesteld worden, dat de diagnose die bij opname gesteld wordt, weinig verband vertoont met het behandelingsresultaat zoals dit door de patiënt zelf ervaren wordt. De psychotherapeut ziet wel verschillen, maar contaminatie mag niet worden uitgesloten. Overigens is de groep patiënten met een diagnose persoonlijkheidsstoornis, welke volgens de therapeut minder effectief is behandeld, gering.

9.5.3 Ontslagjaar.

In hoofdstuk 8 is herhaaldelijk aangetoond, dat er een trend was in het behandelingsresultaat: gedurende de opeenvolgende jaren nam dit geleidelijk af, met een uitzondering in 1984. Over 1985 zijn slechts een beperkt aantal gegevens beschikbaar, zodat deze hier achterwege blijven.

Systematische analyse bij de opeenvolgende vragenlijsten levert weinig statistische significantie voor deze trend; alleen bij de antwoorden van de psychotherapeuten blijkt de factor: beëindiging behandeling tot significante verschillen tussen de jaren aanleiding te geven ($F = 3.27$ $p = .02$). De ombuiging van de trend in 1984 lijkt debet te zijn aan dit gebrek aan significantie. Over de periode 1976-1980 wordt wel een significant verschil in effect signaleerd bij de ontslagvragenlijst van de patiënt ($F = 2.85$ $p = .04$).

Van deze drie hoofdeffecten: therapeut, diagnose en ontslagjaar zijn ook duplex- en triplex-interacties berekend. Dit levert incidenteel significante verschillen op, maar de aantallen per cel zijn dan zo gering dat de resultaten niet meer betrouwbaar geïnterpreteerd kunnen worden.

9.5.4 Andere effecten.

Naast deze drie effecten is vervolgens nagegaan of andere patiënt-kenmerken tot verschillen in resultaat geleid hebben. Hiertoe zijn de variabelen: geslacht, burgerlijke staat, opleiding, inkomen en woonsituatie als onafhankelijke variabelen gehanteerd. De variabele "inkomen" werd gerecodeerd in: werkzaam, ziek/wao, ww/rww/bijstand, studierend/gepensioneerd en geen eigen inkomen (huisvrouw/-man). De opleiding in lager onderwijs, LBO/MAVO, MBO/HAVO/VWO en HBO/WO. Als therapieresultaat is gekozen voor de factor effect van de vragenlijsten. Uit eenzijdige variantie analyse blijkt, dat er alleen een significant verschil is bij de variabele geslacht: mannen geven een iets grotere effectiviteit aan dan vrouwen ($F = 5.05$ $p = .025$), zowel bij ontslag als bij de eerste follow-up. Tenslotte zijn nog als onafhankelijke variabele gekozen de opnameduur en de tijd die verstreek tussen ontslag en start van de voortgezette hulpverlening. De opnameduur laat geen verband zien met het behandelingsresultaat, hetgeen overeenstemt met de in hoofdstuk 2 beschreven verwachting, dat de behandeling in principe duurt totdat een bepaald resultaat bereikt is en er dus geen relatie zal zijn tussen duur en effect (zie ook Hesselink, 1986). Er wordt wel een significant verband gevonden tussen het effect en het starten van de nazorg na ontslag: zowel bij ontslag als bij de eerste follow-up blijkt een significant verschil in gemiddelde verbetering tussen de patiënten die snel (binnen 1 maand) na ontslag voortgezette hulp kregen en zij die hier langer op moeten wachten. Zoals verwacht mag worden, leidt een voortgezette hulpverlening die snel aansluit tot een beter resul-

taat. Omgekeerd laat deze uitslag ook zien dat bij patiënten die minder verbeterd worden ontslagen, later voortgezette hulp start. Dit zal vaak verband houden met een "late drop-out"; wanneer een patiënt (plotseling) eenzijdig de behandeling afbreekt, is doorgaans niet snel voortgezette hulpverlening te regelen.

De laatste variabelen die een relatie zouden kunnen hebben met het behandelingsresultaat, zijn de bij het statistiekformulier genoteerde gegevens: duur van de klachten, duur en mate van mogelijk dysfunctioneren en duur van de voorafgaande behandeling. Geen enkele van deze variabelen correleren met het effect bij ontslag. Bij de eerste en tweede follow-up zijn deze correlaties rond de 0.20, wat bij ca. 300 personen significant is ($p < .001$), maar nauwelijks te interpreteren. De in hfst. 7 gestelde vraag of de gesignaleerde veranderingen in de patiëntenpopulatie consequenties zou kunnen hebben voor het behandelingsresultaat is hiermee voor zover met dit gegevensbestand mogelijk is, beantwoord. Afgezien van het geslacht is er geen relatie tussen deze variabelen en het effect, althans wanneer op deze klassieke (normatieve) wijze wordt getoetst. In hfst. 2 is evenwel betoogd dat een absolute, criterium-georiënteerde toetsing meer is geïndiceerd. Deze wordt in de volgende paragraaf gegeven.

9.6 NIET-EFFECTIEVE PATIENTEN.

In de voorafgaande paragraaf is gezocht naar variabelen die samenhang vertonen met het behandelingseffect. Tenslotte zijn enkele patiëntkenmerken met dit effect gecorreleerd. Een andere benaderingswijze is om die groep patiënten die een te gering resultaat bereiken, nader te onderzoeken. Hiervoor is het noodzakelijk het criterium te bepalen voor een voldoende resultaat. Gelet op de voorafgaande analyses is als operationalisatie voor behandelingseffect gekozen: de gemiddelde verbetering op de door de patiënt geformuleerde doelen bij opname. De norm wel-niet 'effectief' is gelegd bij een score lager dan 3 op een 6-punts schaal, welk punt omschreven is als "min of meer verbeterd". Een dergelijke norm wordt ook gehanteerd door Malan (1976) die op een schaal van 0 (geen verandering) tot 4 (geheel verbeterd) het criterium voor verbetering groter dan gelijk aan 2 legde (zie Schagen, 1983, p96).

Over de totale groep van 318 patiënten voldoet 66% aan dit effectiviteitscriterium. Dit getal heeft in het psychotherapie-effect-onderzoek een bijna magische betekenis sinds de geruchtmakende publicaties van Eysenck, die eveneens in de door hem bestudeerde therapieën een verbetering vond van tweederde van alle patiënten, ongeacht de verkregen behandeling.

Met behulp van een discriminant-analyse is nagegaan of effectiviteit te voorspellen zou kunnen zijn op grond van tal

van patiëntenkenmerken. Middels de berekende discriminantfunctie blijkt slechts 60% juist te classificeren te zijn, hetgeen nauwelijks afwijkt van een toevalstoekenning.

Toch blijft het intrigeren welke variabelen het behandelingsresultaat mede bepalen, of - met andere woorden - welke kenmerken hebben de niet-effectief behandelde patiënten? Deze vraag is daarom van belang, omdat een beleidsverandering zich op deze groep zal moeten richten. Hoe negatief dit ook klinkt, maar ook hier zal men van de fouten moeten leren. Een profiel van de niet-effectieve patiënt kan hierbij behulpzaam zijn. Daarom wordt van enkele nominale variabelen aangegeven wat het percentage niet-effectieve patiënten per categorie is. Het totale percentage niet effectieve patiënten is 34%.

- Psychotherapeut: het percentage niet-effectieve patiënten varieert van 18% tot 47%. Een opmerkelijk groot en significant verschil. ($\chi^2 = 18.89$ df = 5 p = .002)
- Diagnose. Per diagnosegroep is de niet-effectiviteit: fobie = 32%, dwang = 32%, overige = 33%, angst = 34%, depressie = 42% en persoonlijkheidsstoornissen = 47% (niet sign.) Deze laatste categorie bevat slechts 15 patiënten.
- Geslacht. Bij mannen is het effectiviteitspercentage: 69% en bij vrouwen 64%. Dit verschil is niet significant.
- Ontslagjaar. Over de periode 1981 - 1984 is het percentage niet effectief behandelde patiënten resp. 37%, 33%, 46% en 25%. (niet sign.)
- Opnameduur: korte opnames (minder dan 2 maanden) en opnames van 6-9 maanden zijn minder effectief (40% niet effectief) dan de overige perioden (30%) (niet sign.).
- Opleiding: patiënten met de hoogste opleidingen blijken minder effectief (44%) dan de laagste opleidingscategorie (20%) (niet sign.).
- Inkomen: er zijn nauwelijks verschillen tussen de diverse groepen. Per categorie zijn de percentages niet-effectiviteit: werkzaam = 39%, ziektewet/WAO = 34%, WW/bijstand = 35%, studierend = 20%, huisvrouw = 33% (niet sign.).
- Intake: van de drie intakeers, die reeds jaren deze functie vervullen, varieert het percentage niet-effectieve patiënten sterk: 44%, 38% en 28%. Wordt de eerste intakeer die 40% van alle intakes heeft verzorgd, vergeleken met de twee anderen, dan is het verschil in effectiviteit van de patiënten significant ($\chi^2=4.98$ df=1 p=.02)
- Voortgezette hulpverlening: patiënten die bij ontslag nog geen hulpverlener hebben, voor wie de therapeut laat nazorg zoekt - en die bijgevolg vaak lang op hulp moeten wachten - zijn minder effectief behandeld dan anderen. Wanneer snel hulp start (binnen een week), is de niet-effectiviteit 27%. Wanneer voortgezette hulpverlening later dan een maand na ontslag van start gaat, is deze 56%. ($\chi^2 = 13.66$ df = 3 p = .003)

Tenslotte is nagegaan hoe representatief de groep respondenten is, met andere woorden of effectieve patiënten vaker

een vragenlijst retourneren dan niet-effectieve patiënten. Aan de patiënten die minder dan 4 weken opgenomen zijn geweest, is nooit een lijst verzonden. Dit zijn over het algemeen niet-effectieve patiënten (drop outs). Voor de overige 317 patiënten is wel na te gaan of er een relatie bestaat tussen respons en effect. Deze blijkt er niet te zijn. Bij alle ontslag en follow-up vragenlijsten reageert ca. 65% van de effectieve en 35% van de niet-effectieve patiënten. Het toenemende aantal non-responders bij de derde follow-up is eveneens evenredig verdeeld: 68% effectieve en 32% niet-effectieve patiënten. In dit opzicht is er geen aanwijzing dat de patiënten van wie geen follow-up gegevens verkregen zijn, qua effectiviteit afwijken van degenen die wel reageren. Hier spelen kennelijk andere factoren (meer) een rol.

9.7 CONCLUSIES.

Uit de analyse van de ontwikkelde evaluatie-vragenlijsten blijkt, dat deze een hoge interne consistentie hebben alsmede dat de informatie zelf betrouwbaar is (interbeoordelaars- of test-hertest-betrouwbaarheid). In dat opzicht zouden de vragenlijsten kunnen worden ingekort voorzover ze redundante informatie verschaffen.

Factoranalyses laten een duidelijk en goed interpreteerbaar patroon van factoren zien. Aldus kunnen de ca. 700 variabelen worden gecomprimeerd tot gemiddelde scores op 23 factoren. Wanneer deze scores als afhankelijke variabelen worden gebruikt voor het opsporen van verbanden met andere gegevens, dan blijkt met name een significante relatie met de behandelbaar: sommige psychotherapeuten worden lager beoordeeld qua satisfactie en effect dan anderen. Alle andere onafhankelijke variabelen vertonen nauwelijks of geen significante verschillen.

Als de patiëntengroep gedichotomiseerd wordt in effectieve en niet-effectieve patiënten, dan blijkt eveneens dat er grote verschillen tussen therapeuten gevonden worden, van 18% tot 47% niet-effectief. Alle andere variabelen geven kleinere verschillen te zien. Een opmerkelijk verschijnsel is de relatie tussen effectiviteit en het regelen van voortgezette hulpverlening: bij patiënten die effectief behandeld zijn, start eerder voortgezette hulpverlening dan bij degenen die bij ontslag als niet-effectief beschouwd moeten worden, terwijl deze laatste groep deze toch het meest dringend nodig hebben.

Hoofdstuk 10:

Met verschillende maten gemeten

*Unless we can resolve the criterion problem,
our therapeutic enterprise
will remain built on quicksand.*
Strupp, 1976, p. 8.

10.1 INLEIDING.

In hoofdstuk 4 is een overzicht gegeven van operationalisaties van psychotherapie-effecten. Aan het slot werd geconcludeerd dat onderzoek naar het resultaat van psychotherapie in de praktijk nauw moet aansluiten bij de problemen van de patiënt. Geïndividualiseerde doelevaluaties zijn hiervan een voorbeeld. Metingen van veranderingen op standaardvragenlijsten die inhoudelijk aansluiten bij de problematiek van de behandelde patiëntengroep zijn eveneens geïndiceerd, zeker ter vergelijking met de doelmetingen. Door een vergelijkende studie van diverse effectmetingen in een praktijksetting kan een bijdrage worden geleverd aan het criteriumprobleem bij de evaluatie van psychotherapie.

Een dergelijk onderzoek kan indicaties geven over de mogelijkheden om de opgetreden effecten aan te tonen, en de psychotherapeutische onderneming een stevigere bodem geven dan het door Strupp (1976) genoemde drijfzand. Zoals eveneens in hoofdstuk 4 is gesteld, mag in een programma-evaluatie project ook het bepalen van de tevredenheid van de behandelde patiënten niet ontbreken.

In het Project Overwaal Programma Evaluatie zijn gedurende de periode 1980-1984 deze drie soorten meetinstrumenten naast elkaar gebruikt (zie hoofdstuk 6). De resultaten van doelevaluaties die de patiënt zelf formuleerde ("target symptoms") en de cliënt-satisfactie zijn beschreven in hoofdstuk 8 en 9.

Naast deze metingen welke door de patiënt bij opname, ontslag en follow-up zijn verricht, zijn nog twee andere doelevaluatie procedures gebruikt: Individuele Doel Evaluaties door de behandelende therapeut (I.D.E.) en door een Externe beoordelaar i.c. de onderzoeker (I.D.E.E.). Deze laatste blijft hier buiten beschouwing. In een afzonderlijk onderzoek zal de betrouwbaarheid van deze evaluatie-meting worden nagegaan, waarna de resultaten met de overige effectmetingen in verband kunnen worden gebracht. Een voorlopige conclusie is dat deze I.D.E.E. in hoge mate overeenstemt met de doelmetingen van de patiënt, zowel inhoudelijk als evaluatief.

Resteert een vergelijking tussen: "target complaints" (patiënt), satisfactie (patiënt), diverse psychologische vragenlijsten (patiënt), globaal effect (patiënt en therapeut), I.D.E. (therapeut) en de diagnose (therapeut).

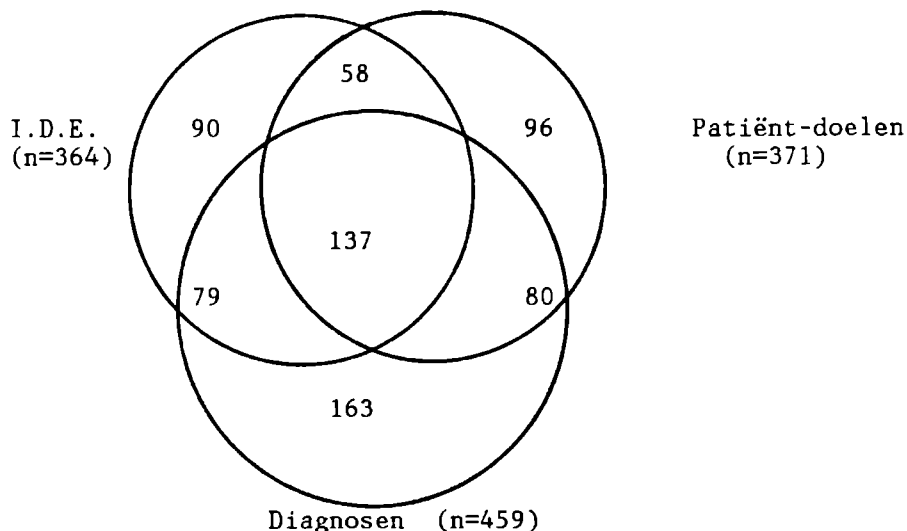
Op drie verschillende manieren is informatie verkregen over de inhoud van de problemen van de patiënt. Door de therapeut werden bij opname diagnoses gegeven volgens de I.C.D.-9. Daarnaast formuleerden patiënt en therapeut behandelingsdoelen, de patiënt middels de "target complaints" en de therapeut in de I.D.E.'s. Deze twee doelmetingen zijn gecodeerd volgens een rubriekenlijst; later zijn de diagnoses hierin eveneens ondergebracht. Uiteraard is dit niet voor alle diagnoses mogelijk. Met name diagnoses als hysterie, persoonlijkheidsstoornissen en psychose komen niet in de rubricering voor, omdat patiënten hiervoor nooit zelf doelen zullen noemen. Deze diagnoses komen resp. 26, 84 en 5 maal voor, in totaal 7% van alle gestelde diagnoses (zie tabel 7.12). Uit onderzoek is bekend, dat doelstellingen van cliënten en hun hulpverleners vaak niet met elkaar in overeenstemming zijn. "De cliënt en zijn hulpverlener, een paar apart" (Hageman-Smit, 1976), was indertijd een geruchtmakende publicatie.

Hier zal worden nagegaan of de inhoud van de doelstellingen met elkaar in overeenstemming zijn en of deze in relatie staan met de gestelde diagnoses. Van 172 patiënten zijn patiënt-doelen, therapeut-doelen en diagnoses bekend. Aangezien het kan voorkomen dat bij één patiënt twee of meer malen hetzelfde doel of diagnose is genoemd, zijn deze dubbele codes verwijderd. Zo resteerden 459 diagnoses, 364 I.D.E.-codes en 371 patiëntdoelen.

Worden nu de doelen van patiënt en therapeut vergeleken, dan blijkt dat 195 maal dezelfde doelen worden geformuleerd, 176 maal noemt een patiënt doelen en de therapeut niet en 169 maal is het omgekeerde het geval. De overeenstemming is daarom ca. 50%. De relatie tussen therapeutdoelen en de diagnoses is als volgt: 216 maal blijken deze inhoudelijk overeen te stemmen, hetgeen ca. 60% van alle doelen is. Hier is de overeenstemming derhalve groter. Worden de doelen van de patiënt vergeleken met de diagnoses van de therapeut, dan is de overeenstemming ook 60%. Kennelijk sluit de abstractere diagnose meer aan bij de doelen van de patiënt dan de doelen van de therapeut dat doen. Deze drie gegevens combinerend, ontstaat de volgende interrelatie: 137 maal stemmen diagnoses, I.D.E. en patiëntdoelen overeen; 79 maal de I.D.E. en de diagnose; 80 maal de patiëntdoelen en de diagnose en 58 maal de I.D.E. en de patiëntdoelen. In figuur 10.1 zijn deze overeenkomsten in een vendiagram weergegeven.

Uit deze relaties kan geconcludeerd worden, dat de inhoudelijke overeenstemming gering is. Opmerkelijk is vooral de matige overeenkomst tussen de doelen van de patiënt en therapeut. De conclusie van het eerder aangehaalde onderzoek van Hageman-Smit wordt hiermee bevestigd.

De onbetrouwbaarheid van de diagnoses was reeds eerder gebleken. Van 1976 tot 1979 gaven therapeuten bij opname (na ca. 1 maand) en bij ontslag afzonderlijk de diagnoses waar-



Figuur 10.1 Overeenkomst tussen doelen van de therapeut, van de patiënt en de diagnosen.

voor de patiënt behandeld werd. Toen werd bij ontslag slechts 58% van de diagnosen genoemd die ook bij opname waren opgegeven. Er werden daarentegen ca. 35% nieuwe diagnosen vermeld. Dit zou kunnen betekenen, dat de bij opname aangegeven diagnosen bij ontslag niet meer aanwezig (over) waren, of achteraf gezien onjuist waren of gewoon een onnauwkeurigheid van de diagnosticus. Om dit na te gaan, werd in 1980 de procedure gewijzigd. Bij ontslag werden de diagnosen die bij opname waren genoemd, weer aan de therapeut voorgelegd. Deze kon nu per diagnose aangeven hoe ernstig deze bij ontslag nog was, of dat deze over was, of achteraf gezien niet van toepassing. Tevens konden nieuwe diagnosen worden opgegeven. Van deze laatste mogelijkheid werd nu weinig gebruik gemaakt, 16% van de diagnosen bij ontslag waren nieuw, terwijl van 17% van de opnamediagnosen bij ontslag werd opgegeven dat deze niet van toepassing of over waren. Men kan zich nu kennelijk veel meer vinden in de diagnosen die eerder waren gesteld. De verschillen voorheen tussen opname en ontslag-diagnose waren kennelijk voornamelijk te wijten aan onbetrouwbaarheid. Deze resultaten herinneren aan het grapje dat de herkomst van het woord "diagnose" zou stammen van het Griekse "δι-αγνωσ", dat "tweemaal onwetend" zou kunnen betekenen (Spitzer, 1980) in plaats van "δι-α-γνωσκειν" hetgeen tweemaal kennen of waarnemen zou betekenen en volgens Kendell zou staan voor "onderscheiden" (Kendell, 1975). Deze afleiding is overigens onjuist. De juiste etymologie van het woord diagnose is "δι-α-γνωσ" dat "door en door weten" betekent, hetgeen zeker niet voor iedere diagnose zal gelden.

Deze onbetrouwbaarheid heeft uiteraard invloed op de inhoudsvaliditeit die hierboven onderzocht is.

10.3 I.D.E. - EEN IDEE FIXE?

Bij de start van het vergelijkende onderzoek van de verschillende doelevaluatie procedures was met de psychotherapeuten overeengekomen, dat zij bij opname c.q. tegelijk met het behandelingsplan ook doelen zouden opstellen volgens de vereenvoudigde Goal Attainment Scaling, Individuele Doel Evaluaties (I.D.E.) genaamd. Bij de verplichte maandelijkse evaluaties zouden deze doelen eveneens gemeten worden. Zo zou een reeks van metingen ontstaan, die de progressie van de patiënt richting doel zou weergeven. Het al dan niet realiseren van het gestelde doel zou vervolgens gerelateerd kunnen worden aan andere uitkomstmaten en patiëntengegevens.

Het invoeren van deze individuele doelevaluaties is het meest moeilijke onderdeel van het evaluatie-project geweest. In de periode februari/maart 1980 werd als proef door iedere therapeut één doel per patiënt geformuleerd. Na enige ervaring met deze manier van doelen formuleren werd geleidelijk ingevoerd om dit voor iedere patiënt te doen en maandelijks te evalueren. Regelmatig was het nodig therapeuten en assistent-psychiaters hiertoe te blijven instrueren en motiveren. Tussen de behandelaars bleken grote verschillen te bestaan in hun houding en bekwaamheid doelen te formuleren, en met name hierbij concrete metingen te construeren.

Na een korte onderbreking vanwege administratieve redenen werden in februari 1981 geautomatiseerde I.D.E.-formulieren ingevoerd (zie ook paragraaf 6.5.3.2). Gedurende ca. 4 jaar werden aldus van 172 patiënten therapeut-doelen verzameld. Dit is ruim 50% van alle patiënten uit die periode. De arts-assistenten bleken zelden doelen te formuleren, alsmede een van de psychotherapeuten. Uit deze gegevens kan al geconcludeerd worden, dat voor deze evaluatieprocedure te weinig animo aanwezig was. Zoals ook door anderen gemeld (Kiesler, 1977 zie ook hfst 4)) moeten de behandelaars bij deze methode permanent begeleid en ondersteund worden, hetgeen een bewijs is dat deze wijze van evalueren weinig kans maakt een routinematige procedure te worden.

Naast deze kwantitatieve gegevens is vervolgens de kwaliteit van de geformuleerde metingen en de evaluatie hiervan beoordeeld. Bij een steekproef van 20 patiënten bleek slechts 35% van de metingen bruikbaar te zijn. De helft van de metingen bevatten uitsluitend subjectieve beoordelingen door de patiënt of de behandelaar en weken nauwelijks af van de reeds op het formulier vermelde globale beoordeling. Soms werd deze zelfs letterlijk overgenomen. Enkele malen werd een meting overdreven ingewikkeld gemaakt. Zo werd een meting voor angst om alleen te zijn aangetroffen, waarbij de tijdsduur dat patiënt alleen in huis was, werd uitgedrukt in $4 + {}^2\log(\text{tijd})$

in uren). De schaal zou moeten lopen van 1 (beginniveau) naar 9 (streefscore). Om de score 1 te bereiken moet de patiënt over een rekenmachine beschikken en 0,125 uur (ofwel 8 min.) alleen thuis zijn. Voor het behalen van de streefscore maar liefst 32 uur. De vraag is of dit door de constructeur bedoeld is of dat hier met de metingen een quasi intellectueel loopje is genomen.

Het regelmatig evalueren van de doelstellingen bleek een andere moeilijkheid. Bij meer dan de helft van de metingen ontbrak een of meer evaluatie-scores. Dikwijls bleek, dat wanneer één evaluatiescore werd gemist, de discipline werd verbroken en er meer evaluaties gingen ontbreken; het evalueren stopte dan doorgaans.

Al deze feiten veroorzaken een gegevensbestand dat zeer onvolledig is van een beperkt aantal psychotherapeuten met te weinig en/of onbruikbare of subjectieve evaluatiegegevens. Om deze redenen zijn deze doelevaluaties niet betrokken bij de evaluatie van de behandeling. De in de vorige paragraaf beschreven vergelijking van de inhouden van de doelen is het enige haalbare op dit moment. Dit neemt evenwel niet weg, dat enkele psychotherapeuten met veel zorg en toewijding voor een groot gedeelte van hun patiënten doelen en metingen hebben geformuleerd en geëvalueerd. Hun werk is dan gedeeltelijk voor niets geweest ten aanzien van dit onderzoek. Deze therapeuten waren evenwel van mening dat dergelijke doelen en evaluaties ook positief werkten voor hun behandeling. Het is dan ook niet verwonderlijk dat doorgaans therapeuten die nauwgezet doelen formuleerden ook een hoger effectiviteitspercentage behaalden.

Gelet op deze praktijkervaring kunnen vraagtekens geplaatst worden bij de mogelijkheid van een dergelijke vorm van behandelingsevaluatie door de psychotherapeuten. Op Overwaal wordt zeer systematisch gewerkt met maandelijke evaluaties van het behandelingsplan. In dit plan staan behandelingsdoelen geformuleerd. De I.D.E.-procedure sluit hier volledig op aan. Daarnaast worden gedragstherapeuten beschouwd als zorgvuldige meters van het behandelingsverloop. Het voorwoord van Behavioral Assessment (Hersen en Bellack, 1976) begint aldus: "One of the hallmarks of behavior modification is its emphasis on empiricism, experimentation, and precision. This emphasis had led to an integral relationship between treatment and assessment. In contrast to most traditional clinical approaches in which assessment is functionally independent of treatment, behavior modification strategy requires careful assessment before, during, and after treatment" (p. IX). Geconcludeerd wordt dat zo dit al voor gedragstherapeuten zou opgaan, de I.D.E.-procedure kennelijk te tijdrovend is en een creativiteit, inventiviteit en discipline vereist, die niet iedere psychotherapeut gedurende lange tijd kan of wil opbrengen.

10.4 DE MATEN VERGELEKEN.

Gedurende de periode 1980-1984 hebben patiënten bij opname, ontslag en follow-up na 6 maanden, diverse standaardvragenlijsten ingevuld. Het is gebruik in het psychotherapie-effectonderzoek de verschillen tussen begin en einde van een behandeling te hanteren als mate van effectiviteit. Als de gemiddelde scores, doorgaans getoetst met behulp van de student-t-toets, significant van elkaar verschillen, is de behandeling effectief geweest.

In deze paragraaf zal deze normatieve evaluatie-procedure vergeleken worden met een criterium-georiënteerde. Ook zal een vergelijking worden gemaakt tussen deze standaardvragenlijsten enerzijds en de doeevaluaties en andere metingen zoals de factoren welke in hoofdstuk 9 beschreven zijn, anderzijds.

10.4.1 De normatieve evaluatie met behulp van standaardlijsten.

Gedurende de jaren 1980, 1981 en 1982 werden bij opname, ontslag en 6 maanden follow-up aan alle patiënten een set vragenlijsten uitgereikt. In 1983 werd de inhoud van dit pakket gewijzigd (zie hoofdstuk 6). De gegevens van de eerste drie jaar zullen hier gebruikt worden. In totaal zijn bij opname 194 volledig ingevulde lijsten terugontvangen, een respons van 95%, exclusief 19 patiënten die voor of tijdens de testprocedure vertrokken. Bij ontslag zijn 168 sets ingevuld. Door een voortijdig vertrek (binnen 1 maand) zijn aan 8 patiënten bij ontslag geen lijsten verstrekt en 21 patiënten vulden om andere redenen geen vragenlijsten in. Deze laatste groep als niet-responders definierend is het antwoordpercentage 89%. Bij follow-up zijn 111 sets verkregen, nog slechts 66% van degenen die bij ontslag meewerkten. De vragenlijsten betreffen: N.P.V., Zung-depressie, M.A.S., F.S.S., IDB, Assertiviteitslijst Brinkman, Klachtenlijst Brinkman, E.P.I., I.W. Hermans.

De gemiddelde scores tussen opname en ontslag, en opname en follow-up zijn op statistisch verschil getoetst met behulp van de t-toets. In tabel 10.1 en 10.2 staan de resultaten vermeld. Afgezien van de scores op de F.S.S. subschalen 3, 4, 5 en 6 zijn alle verschillen tussen opname en ontslag significant ($p < .001$). Tussen opname en follow-up blijken eveneens bijna alle scores significant te verschillen. De meeste schalen van de N.V.P. geven nu geen significant verschil meer aan.

Bijgevolg zou nu geconcludeerd kunnen worden, dat de behandeling dus effectief is geweest voor deze patiënten op alle gebieden die door deze vragenlijsten gemeten zijn, en dat dit resultaat ook nog bestaat een half jaar na ontslag, behalve voor enkele van de meer algemenere schalen van de N.P.V. Met een dergelijke conclusie eindigt dan ook het

Tabel 10.1 Verschil tussen voormeting en nameting per score op de standaardvragenlijsten.

test	schaal	N	voormeting		nameting		toetsing	
			X	Sd	X	Sd	t	p
N.P.V.	IN	164	29.2	8.4	23.4	11.5	7.69	.000
	SI	161	17.9	7.5	15.9	7.6	4.07	.000
	RG	166	27.2	8.7	23.8	8.8	6.42	.000
	VE	165	20.5	7.7	18.5	8.3	4.21	.000
	ZE	165	9.8	4.4	9.1	4.6	2.55	.012
	DO	163	11.4	5.8	12.4	5.8	-2.84	.005
	ZW	166	16.9	7.1	18.7	7.8	-3.78	.000
Zung		166	55.3	9.4	47.5	11.7	9.07	.000
M.A.S.		166	37.7	8.4	33.5	10.2	6.62	.000
F.S.S.tot		166	102.3	48.8	92.4	47.1	3.58	.000
	1.	166	29.0	12.7	24.9	12.6	4.90	.000
	2.	158	17.8	11.8	15.0	11.0	4.18	.000
	3.	164	24.1	13.6	22.7	11.9	1.98	.050
	4.	163	16.9	10.1	15.9	9.8	1.80	.074
	5.	143	11.1	7.5	10.7	7.4	0.89	.377
	6.	145	6.8	4.5	6.6	4.3	0.68	.498
I.D.B.		166	84.1	21.9	76.7	22.3	5.31	.000
Ass.	BOS	166	38.1	13.1	43.3	12.0	-6.67	.000
	AAS	166	28.8	10.9	32.7	10.9	-5.86	.000
	WEI	166	44.3	11.4	49.0	9.2	-6.33	.000
	EMO	166	39.7	10.4	43.6	11.0	-5.04	.000
	SAP	166	35.7	10.9	39.9	11.1	-5.45	.000
	JTD	166	42.9	13.0	47.4	10.9	-5.19	.000
	Tot	166	229.5	55.5	255.8	55.2	-7.51	.000
S.A.		166	62.7	19.0	57.4	19.9	4.84	.000
I.W.	pos.	161	7.2	3.7	8.5	4.4	-4.26	.000
	neg.	165	13.6	3.6	11.6	4.8	5.99	.000
Klachten		166	94.5	24.3	84.3	28.7	5.65	.000
E.P.I.	N	47	13.0	2.8	12.0	2.9	3.02	.004
	E	47	8.7	2.3	10.7	3.1	-4.89	.000

*) p .000 betekent $p < .001$

meeste therapie-effectonderzoek. Hoe terecht dit is, zal uit paragraaf 10.4.3 blijken.

10.4.2 Inhoudsvalidering.

Het overall-resultaat dat alle patiënten tussen voor- en nameting significant verbeteren, is wellicht nog te nuanceren. Immers, enkele F.S.S.-scores geven geen verschil te zien. Het betreft hier hele specifieke symptomen, zoals angst voor sex/agressie, ziekte/wonden/dood, dieren en geluiden. De groep in zijn geheel geeft deze angsten nauwelijks aan.

Tabel 10.2 Verschillen tussen voor- en follow-up meting per score op de standaardvragenlijsten.

test	schaal	N	voormeting		f.u.-meting		toetsing	
			X	Sd	X	Sd	t	p*)
N.P.V.	IN	108	28.6	8.8	22.8	11.7	5.46	.000
	SI	105	17.8	7.8	16.4	8.3	1.91	.059
	RG	108	26.9	8.6	23.7	8.9	4.18	.000
	VE	107	20.5	7.9	19.2	8.2	1.92	.058
	ZE	108	9.6	4.5	9.5	4.8	0.28	.779
	DO	105	11.3	5.8	11.4	6.3	-0.06	.953
	ZW	107	16.4	7.3	18.6	8.5	-3.12	.002
Zung		108	54.9	9.5	46.0	12.9	7.24	.000
M.A.S.		108	37.5	8.5	32.3	11.6	5.36	.000
F.S.S.tot		108	100.3	50.3	87.4	53.0	3.22	.002
	1.	108	28.0	13.2	23.5	14.3	3.76	.000
	2.	101	17.7	11.4	14.0	10.7	3.77	.000
	3.	107	23.9	13.7	21.3	12.5	2.73	.007
	4.	104	16.8	10.4	15.3	11.1	1.73	.087
	5.	89	11.6	7.7	11.2	8.1	0.62	.536
	6.	88	7.0	4.4	6.8	4.3	0.49	.628
I.D.B.		108	84.6	22.9	75.9	25.6	4.61	.000
Ass.BOS		108	37.5	13.7	44.0	13.7	-5.41	.000
	AAS	108	28.9	11.1	33.9	12.8	-4.42	.000
	WEI	108	44.0	12.2	49.0	10.9	-4.77	.000
	EMO	108	40.8	10.8	44.0	12.5	-3.01	.003
	SAP	108	35.4	10.7	41.1	12.8	-4.98	.000
	ITD	108	42.5	13.2	47.4	12.2	-3.99	.000
	tot	108	229.2	58.4	259.4	66.2	-5.38	.000
S.A.		108	62.5	19.2	55.1	21.8	4.01	.000
I.W. pos.		107	7.3	3.5	8.6	4.9	-3.28	.001
	neg.	106	13.5	3.5	11.0	4.9	5.56	.000
Klachten		108	94.8	24.6	82.7	31.7	4.53	.000
E.P.I.N		32	13.5	2.9	12.3	3.7	2.26	.031
	E	32	8.5	1.9	10.1	3.5	-3.08	.004

*) p .000 betekent $p < .001$

Wanneer de totale groep patiënten nu zou worden opgesplitst in deelgroepen die elk een bepaalde diagnose gemeen hebben, dan zou bij een dergelijke groep grotere verschillen gevonden kunnen worden tussen bepaalde metingen dan bij de andere patiëntengroepen? Zo zou bijv. verwacht mogen worden, dat patiënten met een diagnose dwang op de dwangvragenlijst I.D.B. een groter verschil te zien geven dan de totale groep patiënten. Om deze specificiteit van de metingen na te kunnen gaan, worden op basis van de diagnose de patiënten in de volgende groepen ingedeeld: Dwang, Angst/Fobie, Depressie, Persoonlijkheidsstoornissen en Sociale problemen. Als een patiënt een van deze diagnoses heeft, wordt

hij in de bijbehorende categorie geplaatst. Het is mogelijk, dat iemand in meerdere categorieën voorkomt, indien hij verschillende diagnoses heeft gekregen. In tabel 10.3 staan de resultaten vermeld van deze groepen bij 5 specifieke vragenlijsten: I.D.B. voor dwang, M.A.S. voor globale angst, F.S.S. voor fobieën, S.A. voor Sociale angst en de Zung voor depressie. De diagnosegroepen zijn: Dwang (N = 49), Angst/Fobie (N = 104), Depressie (N = 45), Persoonlijkheidsstoornis (N = 28) en Sociale problemen (N = 84).

Tabel 10.3 Verschillen per diagnosegroep gemeten met 5 specifieke vragenlijsten.

Diagnose	opname X1	ontslag X2	test t	sign p	ver- schil	sd.
test: I.D.B.						
Dwang	100.7	88.3	4.25	.000	12.3	20.3
Angst/Fobie	84.0	76.4	4.27	.000	7.5	18.0
Depressie	81.2	76.0	1.73	.09	5.2	20.0
Pers.stoornis	91.4	85.6	1.88	.07	5.8	16.3
Sociale problemen	84.2	77.8	3.07	.003	6.4	19.0
test: F.S.S.						
Dwang	106.3	96.0	2.05	.046	10.3	35.3
Angst/Fobie	110.3	99.6	3.23	.002	10.6	33.5
Depressie	103.7	95.9	1.31	.196	7.8	39.8
Pers.stoornis	117.8	108.8	1.22	.233	8.4	36.3
Sociale problemen	107.7	96.7	2.67	.009	11.0	37.8
test: M.A.S.						
Dwang	36.6	33.1	2.82	.007	3.4	8.5
Angst/Fobie	38.7	34.6	5.29	.000	4.1	7.8
Depressie	39.1	33.9	3.57	.001	5.2	9.9
Pers.stoornis	38.4	34.3	2.58	.016	4.2	8.6
Sociale problemen	38.2	33.5	5.47	.000	4.6	7.7
test: Zung						
Dwang	53.3	45.1	4.74	.000	8.3	12.2
Angst/Fobie	55.9	47.7	7.50	.000	8.2	11.1
Depressie	58.0	47.2	6.07	.000	10.8	12.0
Pers.stoornis	57.5	50.7	3.2	.003	6.8	11.2
Sociale problemen	57.0	48.2	6.79	.000	8.9	12.0
test: S.A.						
Dwang	61.1	56.7	2.34	.024	4.4	13.1
Angst/Fobie	66.1	59.2	4.98	.000	6.8	14.0
Depressie	63.7	57.3	2.55	.014	6.4	17.0
Pers.stoornis	67.7	61.4	2.83	.009	6.3	11.8
Sociale problemen	64.6	60.1	2.85	.006	4.5	14.3

Uit deze tabel blijkt, dat de I.D.B. inderdaad bij de groep met de diagnose dwang het grootste verschil te zien geeft tussen opname en ontslag, de F.S.S. bij de angst-groep en de Zung bij patiënten met een depressie als diagnose. Toch worden ook vaak significante verschillen gevonden bij diagnosen die niet tot het meetbereik van de betreffende lijst behoren. Zo blijkt de Zung altijd een significant verschil te geven, evenals de M.A.S. De F.S.S. is in dat opzicht specifiek: depressies en persoonlijkheidsstoornissen geven bij deze lijst geen significante verschillen.

Voor dit verschijnsel kunnen drie verklaringen mogelijk zijn:

1. de vragenlijsten zijn inderdaad niet zo specifiek in het meten van het bedoelde construct;
2. de constructen zelf zijn nauw aan elkaar gerelateerd: iemand met dwang is ook dikwijls depressief of ernstige fobische problemen zijn verbonden met sociale problemen;
3. het stellen van de diagnose is niet erg valide, bijvoorbeeld door een lage betrouwbaarheid.

Aan de derde mogelijkheid is reeds in paragraaf 10.2 aandacht besteed, en er is in dit opzicht verder geen uitsluitel te geven. De inhoudelijke verbondenheid zou tot uitdrukking kunnen komen door het toekennen van diagnosen uit de verschillende groepen. Dat dit frequent gebeurt, blijkt uit de aantallen per groep: gemiddeld komt iedere patiënt 1,9 maal voor.

Wanneer de door de patiënt geformuleerde doelen als indelingscriterium worden gebruikt, kan worden nagegaan of deze doelen beter aansluiten bij de scores van de vragenlijsten. Hier is dezelfde procedure gevolgd als bij de diagnosen van de therapeut, met dien verstande dat de rubriek persoonlijkheidsstoornissen niet voorkomt. Dit resulteert in de volgende groepen: Dwang (N = 45), Angst/Fobie (N = 117), Depressie (N = 21), Sociale problemen (N = 18). Uit deze aantallen blijkt, dat de overlap minder groot is dan tussen de diagnosegroepen, hetgeen de specificiteit ten goede zal komen. Uit de t-toetsen in tabel 10.4 blijkt dit ook. Patiënten met een doel dat betrekking heeft op dwangproblemen veranderen het meest op de I.D.B. Hetzelfde geldt voor angst/fobie-patiënten en de F.S.S. en M.A.S. en S.A.-score. Depressiepatiënten geven een verslechtering aan van de I.D.B. en F.S.S.-score. Maar zij blijken ook niet significant te verbeteren volgens de Zung-score, terwijl alle andere groepen dat juist wel doen. Deze patiënten hebben wel de hoogste Zung-score bij opname. Uit deze resultaten kan geconcludeerd worden, dat de gebruikte vragenlijsten een nauwe verwantschap vertonen met de door de patiënt geformuleerde doelen. Bij opname blijkt een bepaalde categorie op de inhoudelijk overeenkomstige lijsten de hoogste score te hebben en ook op die lijsten doorgaans de grootste verandering te bereiken vergeleken met patiënten met doelen uit andere rubrieken. Deze relatie bete-

kent dat de doelen en deze standaardmetingen aldus inhoudelijk gevalideerd zijn. Het verband tussen diagnoses en de psychologische vragenlijsten is minder sterk aanwezig.

Tabel 10.4 Verschillen per doelgroep op vijf vragenlijsten.

Diagnose	opname X	ontslag X	t	-----verschil----- p d sd		
test: I.D.B.						
dwang	100.5	86.1	5.28	.000	14.4	18.3
angst/fobie	86.3	78.3	4.94	.000	7.9	17.4
depressie	74.9	79.3	-1.15	.265	-4.4	17.5
sociale problemen	82.6	75.6	2.93	.005	7.0	17.3
test: F.S.S.						
dwang	104.3	93.1	2.17	.036	11.2	34.8
angst/fobie	107.2	95.4	3.42	.001	11.9	37.4
depressie	93.6	96.9	-0.40	.697	-3.2	37.5
sociale problemen	101.9	87.3	2.68	.01	14.6	39.7
test: M.A.S.						
dwang	37.4	34.1	2.98	.005	3.3	7.3
angst/fobie	39.0	34.8	6.02	.000	4.2	7.6
depressie	38.5	36.9	1.06	.304	1.6	7.0
sociale problemen	38.9	33.9	4.52	.000	5.0	8.0
test: Zung						
dwang	54.0	44.5	5.70	.000	9.5	11.2
angst/fobie	55.6	48.0	7.52	.000	7.6	10.9
depressie	58.4	52.8	1.73	.100	5.6	14.9
sociale probleme	57.4	47.4	6.69	.000	10.0	10.9
test: S.A.						
dwang	61.2	56.5	2.38	.022	4.7	13.2
angst/fobie	65.3	59.0	4.50	.000	6.3	15.1
depressie	60.0	57.4	0.65	.525	2.5	17.9
sociale probleme	63.5	56.6	3.22	.002	7.0	15.7

10.4.3 Criterium georiënteerde evaluatie.

De gebruikelijke en in paragraaf 10.4.1 gepresenteerde evaluatie van therapie-effecten is gebaseerd op statistische significantie. Het bezwaar hiervan is, dat de gevonden verschillen klinisch van weinig betekenis kunnen zijn. Bij grote groepen met een beperkte variantie kunnen kleine verschillen reeds tot significante resultaten leiden. Een versterkend effect is de regressie naar het gemiddelde bij de tweede

meting, aangezien de eerste meting doorgaans extreem hoog zal zijn.

Van de hiervoor gebruikte vragenlijsten zijn ook zogenaamde normaalwaarden bekend. Deze normaalwaarden, berekend op grotere groepen personen zonder (bekende) problematiek, worden uitgedrukt in afwijkingen van het gemiddelde. Dikwijls zijn de schalen eenzijdig: laag is normaal en zeer hoog is afwijkend c.q. de betrokkene heeft in ernstige mate last van hetgeen de lijst meet. Het ligt in een criterium georiënteerde situatie voor de hand deze normaalwaarden te hanteren als criterium voor verbetering: hoeveel patiënten voldoen bij opname niet aan deze norm en hoeveel bij ontslag. Voor de lijsten Zung, M.A.S., F.S.S., I.D.B., A.S.S. en S.A. worden dergelijke vergelijkingen uitgevoerd, zowel voor de totale groep als per doelgroep. In tabel 10.5 staan de verdeling van de normscores vermeld.

Tabel 10.5 Normscores van de gebruikte vragenlijsten.

	Zung	MAS	FSS	IDB	SA	sub-ass.
laag	-	<18	-	-	-	-
normaal	<40	18-26	<75	<83	<56	>248
licht	40-47	-	75-118	83-97	-	-
matig/hoog	48-55	27-35	119-162	98-117	56-71	213-249
zeer hoog	>55	>35	>162	>117	>71	<213

De hier gehanteerde norm is de bovengrens van de range van normaal. Door de grote gevoeligheid van de Zung is hier de rubriek licht depressief bij de normaalwaarde gerekend. Over deze norm is uiteraard discussie mogelijk. Bij het ontbreken van een voor deze behandeling gespecificeerde maatstaf voor de beoogde verbetering, en om de grote discrepantie tussen de statistische toetsing en de klinische relevantie zichtbaar te maken, is voor deze normgrens van "hoog normaal" gekozen. Overigens is deze vergelijking ook elders gebruikt (Ciarlo, 1976, 1978) Er kunnen nu vier mogelijkheden optreden: bij opname en ontslag zijn beide scores normaal, ze zijn beide hoog, bij opname zijn ze normaal en bij ontslag hoog of - de beoogde verandering - bij opname hoog en bij ontslag normaal. In tabel 10.6 zijn de procentuele resultaten van deze vier mogelijkheden weergegeven.

Uit deze resultaten blijkt, dat het percentage patiënten dat onveranderlijk hoog scoort, steeds het hoogste is, behalve bij de I.D.B.. Bij deze lijst is echter het percentage dat onveranderlijk laag scoort het hoogst: 48%. Bij de helft van de patiënten hebben geen dwangproblemen. Wordt het effect nu afgemeten aan de verandering van hoog naar (hoog)

Tabel 10.6 Veranderingen tussen opname- en ontslagscore criterium georiënteerd weergegeven (in %).

Vragenlijst	N-N	H-N	N-H	H-H
Zung	14	31	2	53
M.A.S.	9	14	1	76
F.S.S.	27	16	8	49
I.D.B.	48	21	4	27
S.A.	28	17	7	48
Assertiviteit	28	28	8	36

N = bovengrens van normaal, H = hoger dan normaal

normaal, dan blijkt deze slechts bij 14% - 31% wordt gerealiseerd. Hiermee wordt eens te meer aangetoond, dat criterium georiënteerde evaluatie een beter inzicht geeft in de werkelijke veranderingen dan de relatieve statistische toetsing.

Bij opsplitsing in doelgroepen blijkt, dat van de patiënten met een doel op het gebied van angstproblemen bij ontslag nog 85% een hoge M.A.S.-score heeft (was 96% bij opname), 56% een hoge Zung-score (was 83%), 60% een hoge F.S.S.-score (was 68%), 47% een hoge assertiviteitsscore (was 66%) en 59% een hoge S.A.-score (was 67%). Bij de patiënten met dwang was de I.D.B.-score bij opname bij 80% hoog, bij ontslag nog bij 44%. Bij de depressie-patiënten had 100% een hoge Zungscore bij opname en nog 69% bij ontslag.

Hoewel de veranderingen tussen opname en ontslag niet groot zijn, blijkt deze voor de patiënten met een dwangprobleem het grootst. Hiermee is tevens aangetoond, dat een gedifferentieerde evaluatie noodzakelijk is. Zelfs in een betrekkelijk homogene groep als neurotische patiënten blijken er verschillen per subgroep op basis van hun problematiek te bestaan.

10.4.4 Intercorrelaties effectmetingen.

Tot slot rest de interrelatie tussen de diverse effectmetingen die naast elkaar zijn gebruikt. Van de psychologische vragenlijsten kunnen de verschillen tussen opname en ontslag als effect van de behandeling gezien worden, naast de verbeteringsscores van de patiënt bij ontslag op de geformuleerde doelen en de 23 factoren welke uit de eigen vragenlijsten gedestilleerd zijn.

Worden de verschillen gecorreleerd met de factoren uit de vragenlijsten, dan worden de volgende relaties zichtbaar:

- alle verschillen correleren significant ($p < .01$) met de gemiddelde verbeteringsscores van de doelen van de patiënt bij ontslag ($.35 < r < .50$, $N = \text{ca. } 150$) en de globale beoordeling van het effect ($.20 < r < .40$);

- er is geen verband tussen deze verschillcores en de satisfactie van de patiënt bij ontslagover de totale behandeling en over de psycho- en sociotherapeut;
- de beoordeling van het effect door de psychotherapeut correleert alleen significant met de verschillcores van de M.A.S. en Assertiviteitslijst;
- de globale effectcores van de drie follow-up lijsten van de patiënt correleren met de verschillcores van de Zung, M.A.S., F.S.S. en Assertiviteitslijst;
- er zijn, op enkele na, geen significante correlaties met de factoren van de follow-up vragenlijsten die de voortgezette hulpverlener heeft ingevuld.

Wordt ook nu weer de totale groep opgesplitst in doelgroepen op basis van de patiëntdoelen, dan blijken de correlaties tussen de verschillscore van een bepaalde lijst en de verbeteringsscore op het bijbehorende doel te stijgen. Zo is de correlatie tussen de doelscore depressie en de verschillscore van de Zung: $-.70$; wordt de I.D.B.-vershillscore gecorreleerd met een doelscore die betrekking heeft op dwangproblematiek, dan stijgt deze naar $-.44$.

Deze correlaties bevestigen de voorafgaande resultaten: een evaluatie over de totale groep betekent een nivellering van effecten. De relatie tussen doelmetingen en globale beoordeling van het effect door de patiënt en de psychologische vragenlijsten is duidelijk aanwezig, terwijl de overige beoordelingen nauwelijks of niet hiermee correleren.

10.5 CONCLUSIE.

In dit hoofdstuk zijn de verschillende effectmetingen met elkaar vergeleken. Doelmetingen van de psychotherapeut blijken in de praktijk niet bruikbaar te zijn. Er worden te veel gegevens gemist en de inhoud van de metingen is vaak te subjectief. Sommige therapeuten bleken hoegenaamd niet bereid of in staat dergelijke Individuele Doel Evaluaties te construeren. Deze evaluatiemethode lijkt in een praktijksituatie ongeschikt.

Een tweede conclusie is, dat diagnosen, therapeutdoelen en patiëntdoelen inhoudelijk weinig overeenstemming vertonen. De patiëntdoelen blijken wel sterk gerelateerd te zijn aan andere gegevens. Deze doelen hebben hierdoor een grote bruikbaarheidswaarde.

Evaluatie met behulp van standaardvragenlijsten is de meest gebruikte procedure voor psychotherapie-effectonderzoek. Als de verschillen tussen voor- en nameting relatief getoetst worden, kan dit tot onjuiste conclusies leiden. De in dit onderzoek gevonden statistische verschillen komen niet overeen met klinisch relevante veranderingen. Het merendeel van de patiënten met een bepaald probleem heeft dit nog bij ontslag; slechts een kleine groep geeft een beteke-

nisvolle verandering te zien. Tevens kan worden geconcludeerd dat de patiënt een betrouwbare beoordelaar van zijn functioneren is: de informatie die hij middels de verschillende lijsten en procedures verschaft, is zeer consistent. Deze conclusie werd ook door anderen reeds getrokken (zie Koster, 1986).

Tenslotte blijkt uit de intercorrelaties tussen alle gebruikte evaluatiematen, dat effectmetingen onderling hoog correleren en dat deze nauwelijks met andere gegevens als satisfactie e.d. verband houden. Het meten met verschillende maten blijft daarom aanbevolen.

Hoofdstuk 11:

Programma-evaluatie geëvalueerd

*Et conari pulchrum erat.
(Ook het pogen was schoon)*

11.1 INLEIDING.

Een project dat tot doel heeft het functioneren van anderen te evalueren, dient ook zelf geëvalueerd te worden. Echter, zoals therapeuten niet de meest geschikte beoordelaars van hun eigen functioneren blijken te zijn, zal dit ook voor de evaluator gelden, wanneer hij zijn eigen handelen moet beoordelen. Dit zou eigenlijk weer door een onafhankelijke onderzoeker dienen te gebeuren. Maar aangezien ook in dit onderzoek therapeuten als informatiebron zijn geaccepteerd, zal ook de onderzoeker zijn indrukken geven over de betekenis en effecten van zijn functioneren onder conditie van deze subjectiviteit.

Er is nog een tweede handicap: de evaluatie van het programma Overwaal is gebaseerd op ca. 800 patiënten, terwijl de evaluatie van het Project Overwaal Patient/Programma Evaluatie (P.O.P.E.) in feite een N = 1-studie is, en generaliseren vanuit gevalsbeschrijvingen is vol risico's. Ondanks deze bezwaren wordt hier toch een poging gedaan het project als zodanig te evalueren, omdat hierin ervaringen zijn opgedaan die ook voor andere projecten van nut kunnen zijn. Daarnaast is in hoofdstuk 3 gesteld, dat het tevens van belang is kennis te nemen van de relatie tussen evaluator en geëvalueerden of geëvalueerde instelling. Een activiteit die nog zo in de kinderschoenen staat als programma-evaluatie, zal ook van zijn eigen kinderziekten moeten leren. Dit hoofdstuk somt er enkele op, die in dit project - soms na een lange incubatietijd - aan het licht zijn gekomen.

11.2 DE DOELSTELLINGEN.

Aangezien evalueren betekent: *nagaan of de doelstellingen zijn gerealiseerd*, zullen deze hier successievelijk worden besproken. De doelstelling waarmee P.O.P.E. - toen nog de afkorting voor Project Overwaal Patient Evaluatie - in 1976 startte, was betrekkelijk vaag: *het evalueren van het behandelingsproces*. Deze evaluatie zou moeten leiden tot een *verbetering van het behandelingsresultaat*.

In de periode 1976-1980 is hiertoe een evaluatiesysteem opgezet en ingevoerd dat aan het eerste deel van deze doel-

stelling voldeed: er werd periodiek informatie verzameld over het verloop en effect van de behandeling en de periode erna. Deze informatie werd teruggekoppeld naar de behandelaars en leiding van het instituut. Er waren echter geen aanwijzingen dat deze feedback leidde tot een verbetering van het behandelingsresultaat; in hoofdstuk 8.3 is beschreven dat dit resultaat geleidelijk verminderde. Hoewel 1984 weer een positiever resultaat liet zijn, waren effectmetingen in 1985 weer lager.

Tijdens een studiereis over programma-evaluatie in de Verenigde Staten werd duidelijk dat de doelstellingen van de organisatie zelf c.q. de beoogde effectiviteit van het programma (de behandeling) geëxpliciteerd dienden te worden. Dit is in overleg met de chef de clinique en de psychotherapeuten geprobeerd, waarbij het door Walker gehanteerde schema (zie Sijben, 1979a, p. 6) werd gebruikt. Dit leidde echter niet tot het beoogde resultaat en dit overleg verzandde. Het bleek niet mogelijk om deze doelen precies te formuleren.

Vanzelfsprekend is het niet een eenvoudige opgave om op het terrein van de psychotherapie dergelijke institutionele doelen te expliciteren, anders dan die welke betrekking hebben op aantallen patiënten, behandelingsduur e.d. Doelen aangaande het na te streven resultaat van de behandeling zijn moeilijker te formuleren. Echter op Overwaal bestond op dat moment reeds enige jaren informatie over dit behandelingsresultaat, gemeten met instrumenten die in overleg met de medewerkers waren opgesteld en waarvan de resultaten periodiek waren gerapporteerd en o.a. in het jaarverslag waren vermeld. Hieraan hadden de instituutsdoelen kunnen worden ontleend. Het feit dat dit ondanks deze informatie toch niet lukte, kan er op wijzen dat ook of met name andere factoren een rol speelden om het expliciteringsproces niet te voltooien.

Dit had tot gevolg dat de doelen nu *impliciet* bleven en (door de evaluator) ontleend werden aan de evaluatiegegevens zelf: de resultaten van het ene jaar werden vergeleken met die van het (de) voorafgaande jaar (jaren). Een daling werd onwenselijk geacht, maar bij het ontbreken van een *criterium* hoefde dit niet noodzakelijkerwijze tot actie te leiden. Dit betekende dat van de gegeven feedback geen gebruik werd gemaakt. In de volgende paragraaf zal op de achtergronden van deze konstatering worden ingegaan.

In 1980 werd P.O.P.E. - nu staand voor Project Overwaal Programma Evaluatie - uitgebreid met onderzoekstechnische doelstellingen: welke meetinstrumenten en -procedures zijn bruikbaar en toereikend om behandelingsresultaten aan te tonen. De subdoelen luiden:

- *het ontwerpen van een evaluatiesysteem*, zodat voldoende inzicht verkregen wordt over het verloop en effect van het instituutsfunctioneren;
- *het uittesten van dit systeem*. Wat zijn de praktische mogelijkheden van de gekozen meetprocedures. Worden bijv. de therapeutdoelen steeds tijdig geëvalueerd in dusdanige concrete termen dat de evaluatie berust op meer dan louter subjectieve indrukken;

- *wat is de interrelatie tussen de gebruikte meetinstrumenten.* In hoeverre worden bijv. veranderingen in de scores op de doelevaluaties ook gerepresenteerd in de uitslagen van de standaardvragenlijsten;
- *wat zijn de consequenties van een programma-evaluatiesysteem.* Kan de ingewonnen informatie leiden tot een verbetering van het functioneren (feedback). (Sijben 1981 p. 8; 1982, p.2).

Deze subdoelen zijn over het algemeen in de periode 1980-1985 gerealiseerd. Over elk ervan is in het voorgaande uitgebreid verslag gedaan. Samenvattend kan gesteld worden, dat:

- er een (vergelijkend) evaluatiesysteem ontworpen is, dat voldoende inzicht geeft in het effect van het functioneren;
- dit systeem in de praktijk is getest en per instrument ervaring is verkregen over de bruikbaarheid ervan;
- de interrelaties tussen meetinstrumenten aangetoond zijn.

Aangaande de laatste doelstelling betreffende de feedback-functie moet helaas geconcludeerd worden dat deze niet tot een verbetering van de resultaten heeft geleid. Althans niet in de periode dat dit project werd uitgevoerd. Er werd periodiek op diverse niveau's informatie verstrekt, maar er zijn geen aanwijzingen dat deze feedback benut is. Integendeel, het behandelingsresultaat in zijn geheel verminderde, met name in 1983 (zie hoofdstuk 8.4). De stijging van het resultaat in 1984 is in 1985 voor zover bekend weer geheel verdwenen. De periodieke rapportage leidde echter tot een toenemende weerstand tegen deze informatie.

Volgens de rubricering van Larsen & Werner (1981, p. 80) kan de benutting van de evaluatiegegevens van P.O.P.E. geclassificeerd worden in de laagste categorie: *"Non utilization, information has been considered by a potential user but then rejected or nothing is done with the information"*. Deze teleurstellende constatering vraagt om een nadere analyse van de oorzaken hiervan, zeker omdat dit geen eenmalige ervaring is, maar vaker door evaluatoren wordt gemeld.

11.3 PROGRAMMA-EVALUATIE EN ORGANISATIE-VERANDERING.

Het doel van programma-evaluatie is kortgezegd: verbeteringen bewerkstelligen. Het middel dat wordt gebruikt, is feedback. In een eenvoudig cybernetisch model zal feedback die aangeeft dat er een verschil bestaat tussen de norm en de werkelijke situatie ertoe moeten leiden dat het werkproces verandert. Echter, deze voorstelling van zaken, waarop programma-evaluatie is gebaseerd, is te simpel. Op Overwaal is volgens dit model gewerkt, en geconstateerd is dat de informatie niet die invloed heeft gehad die ervan verwacht kon worden. En gelet op de publicaties en waarschuwingen van andere evaluatoren is dit niet een uitsluitend instituutsg-

bonden gegeven. Geconcludeerd zou kunnen worden dat dit feedbackmodel te eenvoudig is en dat er aspecten aan dienen te worden toegevoegd om de werkzaamheid ervan te verhogen c.q. te veralgemeniseren.

In het inleidende hoofdstuk is de vergelijking gemaakt tussen *profit* en *non-profit organisaties*. Gesteld werd dat deze laatste een belangrijk sturend mechanisme onthouden wordt: het resultaat van hun handelen is niet gekoppeld aan beloning: het budget wordt niet (direct) bepaald door de geleverde prestatie. In de profit-sector bestaat wel een directe koppeling tussen prestatie en inkomen. Deze prestatie dient steeds gericht te zijn op *de behoefte van het publiek*, welke behoefte regelmatig verandert, o.a. door de reclame van de concurrent. Als bijv. de samenstelling van een wasmiddel zelf niet verandert, dan zal minstens de verpakking veranderen, al was het maar de opdruk "Vernieuwd". Deze externe druk maakt dat een profit organisatie altijd *veranderingsgericht* moet zijn, wil deze kunnen voortbestaan en daarmee het inkomen en de werkgelegenheid van de medewerkers, verzekerd zijn.

Deze noodzaak tot aanpassen aan de behoefte is in de non-profit sector minder sterk aanwezig en wordt niet gestuurd door marktmechanismen. Haar *monopoliepositie* is hier o.a. debet aan: een G.G.Z.-instelling hoeft niet te adverteren voor klanten, doet geen onderzoek naar de behoefte van het publiek en hoeft niet de concurrent in de gaten te houden; die is er nauwelijks. De enige strijd die geleverd moet worden, is die met de financier over het budget, maar dat wordt nauwelijks bepaald door *de kwaliteit en de effectiviteit van de dienstverlening*. aantal patiënten per jaar of de behandelingsduur zijn parameters die een rol spelen, maar over de kwaliteit van de zorg geven deze geen indicatie. Daarnaast kunnen deze variabelen zelf worden gemanipuleerd. Een dergelijke organisatie is nauwelijks gedwongen zich te richten op verandering. Welke rol moet feedback dan spelen?

Een handicap is uiteraard hoe het resultaat moet worden aangetoond. Therapie-effect is nu eenmaal moeilijker meetbaar dan verkoopcijfers of winstpercentages. Maar deze moeilijkheid mag niet als alibi fungeren om niet te proberen deze resultaten aan te tonen. Het was immers het uitgangspunt van P.O.P.E. dat het behandelingsresultaat aantoonbaar was en het effect zou moeten worden aangetoond. Hiertoe was in de loop der jaren een meetinstrument ontwikkeld, verbeterd en intern geaccepteerd. De "verkoopcijfers" waren dus bekend. Het feit dat deze niet tot (effectieve) beleidswijzigingen aanleiding gaven, had dus andere redenen. Als voor de G.G.Z. als zodanig geen noodzaak bestaat om de effectiviteit te bewaken, was dit dan voor Overwaal anders?

Wordt vervolgens in dit licht het evaluatieproject van Overwaal bezien, dan krijgen enkele gebeurtenissen meer betekenis. Vóór de start van het project in 1976, in feite nog in de sollicitatieprocedure van de onderzoeker, vond op zijn verzoek een gesprek plaats met het team van behandelaars, om deze behoefte te peilen. Toen bleek dat deze in meerderheid

nauwelijks of niet op de hoogte waren van de komst van deze functionaris en wat hij dan wel zou moeten doen. Zij hadden hier niet om gevraagd, maar waren er ook niet op tegen dat hij onderzoek zou verrichten. De behoefte aan feedback bleek in ieder geval niet. De baan werd toch aanvaard, in de (optimistische) veronderstelling dat het team wel van het nut van feedback overtuigd zou raken als deze eenmaal werd gepresenteerd. En wellicht hadden de meeste teamleden dezelfde instelling.

De eerste jaren werden voornamelijk besteed aan het opzetten van het evaluatiesysteem: het ontwikkelen van vragenlijsten en het verzamelen van de gegevens. Geleidelijk werden de resultaten teruggekoppeld, naar de therapeuten, het team en aan de instelling als geheel (jaarverslag). Deze feedback werd in dank aanvaard. De resultaten aanvankelijk waren redelijk goed, er bestond dus geen spanningsveld. De coöperatie van de psychotherapeuten was over het algemeen zeer goed.

Inmiddels was extern wel bekendheid gegeven aan het bestaan van dit evaluatieproject. Overwaal liep in dit opzicht vooruit, in de G.G.Z. werd nog nauwelijks geëvalueerd. De financier (COZ/COTG) zag dit met welgevallen, en de Geneeskundig Inspecteur van de Geestelijke Volksgezondheid te Arnhem vroeg zelfs de interne rapporten op. Deze bevatten overigens slechts globale informatie.

Sinds 1980 werden er jaarlijks *feedbackbesprekingen* gehouden met het team van psychotherapeuten. Men werden resultaten gepresenteerd waar ze niet zelf om hadden gevraagd. Van de dalende effectiviteit werd - door sommigen bezorgd - kennis genomen, maar deze werd niet vertaald in een verandering van het beleid.

Het is uiteraard niet eenvoudig om feedback te vertalen in beleidswijzigingen. Dalende verkoopcijfers als zodanig geven geen indicatie welk product wel gewenst is c.q. wat aan het product dient te veranderen zodat dit meer verkocht wordt. Zij geven vooral aan dat dit moet veranderen. *Intern bedrijfsonderzoek* (kwaliteitsbevordering) of *extern behoefteonderzoek* (veranderende markt) kunnen de noodzakelijke indicatie geven over gewenste beleidsmaatregelen.

De vergelijking tussen verkoopcijfers en evaluatieresultaten kan mank gaan, wanneer de evaluatie-rapportage minder eenduidig is. Deze moet voor de betrokkenen voldoende begrijpelijk zijn en de relatie tussen evaluator en geëvalueerde moet van dien aard zijn, dat deze acceptatie van de gegevens zal bevorderen. Deze vragen zijn door de onderzoeker zelf niet objectief te beantwoorden. Een onafhankelijke onderzoeker zou in een zgn. *benuttingsonderzoek* deze aspecten kunnen nagaan. De indruk bestaat echter dat problemen op dit vlak nauwelijks hebben gespeeld - althans deze zijn nooit geuit tijdens de jaarlijkse feedbackbesprekingen.

Een derde vraag, en zo niet nog belangrijker, is in hoeverre de resultaten überhaupt beleidsrelevant waren. In sociaal-wetenschappelijk onderzoek wordt vaker geopperd dat de gegevens niet bruikbaar c.q. gebruikt zijn. Cronbach

(1980) formuleerde als 9e stelling: *"Commissioners of evaluations complain that the messages from evaluations are not useful, while evaluators complain that the messages are not used"* (p. 3). Ook op Overwaal werd vaker geopperd dat de resultaten niet bruikbaar waren, omdat niet aangegeven werd wat er moest veranderen. Deze vraag heeft direct betrekking op de taak en de doelstelling van programma-evaluatie. Dient zij "slechts" de status quo te beschrijven, zo mogelijk uitgesplitst in subdoelstellingen/-groepen of ook beleidsaanbevelingen te doen. Alle handboeken over programma-evaluatie nemen als standpunt in, dat de evaluator niet meer mag doen dan feedback geven. Het bepalen van (nieuw) beleid is de taak van de opdrachtgever c.q. de geëvalueerde instelling. Beleid wordt immers door meer zaken bepaald dan door effectiviteitscijfers alleen. In hoofdstuk 3 is hieraan uitgebreid aandacht besteed en gemeld dat volgens de richtlijnen voor programma-evaluatie de evaluator onderscheid dient te maken tussen zijn uitsluitend evaluatieve rol en iedere andere rol waarin hij een standpunt inneemt c.q. verdedigt. Daarnaast is de evaluator zelden in de positie om te kunnen beoordelen welke belangen en welke prioriteiten bij dit beleid een rol spelen. Beleidsvoerders zullen overigens zelden toestaan, dat evaluatoren hun taak overnemen. De bruikbaarheid van de informatie kan dan niet direct worden bepaald behalve dan door de mate waarin deze aangeeft of de beoogde doelstellingen wel of niet worden bereikt. Gesteld wordt dat deze informatie per definitie beleidsrelevant is, althans behoort dit te zijn.

Vanuit sociaal psychologisch oogpunt verstoorde deze feedback evenwel een bestaand evenwicht (Broskowski & Driscoll, 1978, p. 53 e.v.), hetgeen uiteraard niet gewenst was. Ook Haley (1975) beschrijft dit reeds: *"...Mental health clinics are not primarily concerned with treatment outcome. If they were, every clinic would have a research unit which examines results and changes therapy procedures on the basis of the findings. Mental health clinics have never changed a therapeutic approach or discharged a psychiatrist, psychologist, or social worker on the basis of follow-up results of therapy. More important than therapy results to any clinic is staff harmony. If staff members are to work together in reasonable tranquility, they must be in agreement about what should be done."* Het feit, dat dit artikel door een vijftal vooraanstaande Amerikaanse tijdschriften is geweigerd, betekent niet dat hij ongelijk heeft, misschien wel integendeel.

Intussen bleef de evaluator de medewerkers middels nadere analyses informatie verschaffen over mogelijke verklaringen van de verminderde effectiviteit. Zo werden de resultaten uitgesplitst volgens diverse variabelen, maar geen enkele liet significante verschillen zien, behalve de therapeut-variabele (zie hoofdstuk 9). Nu ontstond, behalve het gedogen van het onderzoek, weerstand ertegen. Een illustratief voorbeeld is het volgende voorval. De resultaten werden anoniem gepresenteerd. Men kende alleen zijn eigen gegevens. De the-

rapeut met het laagste effectiviteitspercentage openbaarde zijn eigen nummer, en bestreed vervolgens de juistheid van de cijfers (weerstandsreactie nr. 2: ontkennen, zie hoofdstuk 3). Hij meende aan te kunnen tonen dat de cijfers onjuist waren. Bij een bepaalde vraag stond dat 50,1% van zijn patiënten positief oordeelden over het effect, en 49,9% negatief. Dit zou volgens hem, gezien het marginale totaal, niet kunnen kloppen. Met deze persisterende opmerking werd ernstige twijfel gezaaid over de betekenis van de resultaten, hetgeen wellicht ook de bedoeling was. Begrijpelijkwijze werd door hem een "scape goat" positie verworpen. (Bij narekening bleek later dat de gegeven uitslagen inderdaad niet geheel juist waren: er had moeten staan 50,0% en 50,0%.)

Dergelijke reacties zijn organisatie-psychologisch geheel begrijpelijk: er wordt informatie in een systeem teruggevoerd dat deze niet kan hanteren. De fout wordt vervolgens gemaakt van de zgn. "*geïsoleerde individu*"-benadering, "zij biedt maar zelden een deugdelijke definitie van een te veranderen eenheid omdat een te groot aantal beslissende invloeden die het individu in de hand heeft, zijn gedrag bepalen" (zie Blake & Mouton, 1977, p. 230). (Bij overweging van dit citaat meende ik, dat dit onjuist moest zijn; veranderen in: "... die het individu *niet* in de hand heeft", zou dit logischer maken. Raadpleging van de oorspronkelijke Engelse versie: The Managerial Grid (1964) leverde het volgende citaat: "... the individual-in-isolation is, rarely, a valid definition of a change unit because too many critical influences *outside* his control determine his behavior" (p. 260). De foute vertaling heeft vijf herdrukken overleefd). Lewin gaf reeds in 1946 de formule: $G = F(P, O)$; Gedrag (G) is een Functie (F) van de onderlinge afhankelijkheid tussen een Persoon (P) en zijn Omgeving (O). Veranderingen dienen derhalve op de *totale situatie*, persoon én omgeving, gericht te zijn. Het product van een organisatie is immers de verantwoordelijkheid van allen die daarin werkzaam zijn. Samenvattend: als in een instelling geen noodzaak tot verandering bestaat (intrinsiek of extrinsiek) en er geen "Planning of Change" (Bennis e.a. 1969) is, kan feedback geen functie vervullen. Weerstand is dan een natuurlijke reactie.

Geconcludeerd wordt, dat programma-evaluatie alleen effectief kan zijn in organisaties die een attitude tot verandering bezitten, welke attitude zo mogelijk kan worden versterkt door externe druk, zodat deze niet te vrijblijvend of als luxe wordt beschouwd. Programma-evaluatie dient te passen in een *organisatie-ontwikkelingsstrategie*. Zo niet, dan kunnen de resultaten hoogstens voor "public relation"-doeleinden worden gebruikt, zoals Beenackers (1984a, zie ook hoofdstuk 3) en Hofstee (1980, zie ook hoofdstuk 2) reeds betoogden. De door Beenackers gekozen oplossing voor externe evaluatie om zo een grotere druk te kunnen uitoefenen, is vol risico's, het zal ook de weerstand vergroten. Een tactiek die erop gericht is om deze te verlagen, heeft wellicht meer kans van

slagen. Watson (in Bennis e.a. 1969, p. 496 e.v.) somt 12 aanbevelingen op om deze te verminderen. Enkele hiervan zijn in dit project gebruikt, maar kennelijk nog niet voldoende.

Een bijkomend aspect is de plaats van de evaluator in de beslissingshiërarchie. Een van buiten komende evaluator die zelf geen deel uitmaakt van de organisatie, geen lijnfunctie bekleedt, maar als staffunctionaris werkzaam is, heeft geen invloed op het beleid. In 1979 werd dit door Flanagan reeds als manco gezien van de situatie waarin de evaluator op Overwaal zich bevond (zie Sijben, 1979, p. 21). Voor Beenackers in het Crisis Centrum te Utrecht gold hetzelfde, zo niet nog sterker, omdat hij als externe evaluator was aangesteld en van meet af aan een bedreiging was (zie Heys, 1984). Een evaluator die wel een lijnfunctie bekleedt, is evenwel niet zonder meer in het voordeel. Als zijn feedback niet wordt gehonoreerd, is zelfs zijn positie in het geding en wordt zijn plaats in de hiërarchie van de organisatie ondergraven. De positie van lijn- of staf-fuctionaris is wellicht ondergeschikt aan de *attitude* bij de ontvangers van de feedback. Als deze niet veranderingsgericht is, zal een stafpositie verkieslijker zijn, als hij al besluit de programma-evaluatieve taak uit te voeren. Bij de start van een programma-evaluatieproject zal deze attitude zelf gepeild c.q. veranderd moeten worden.

De eerste inspecteurs van het staatstoezicht op krankzinnigen (zie hoofdstuk 1) hadden ook reeds geen vrede met een rapportage die niet werd gebruikt. Zij zijn uiteindelijk toch geslaagd: Voor hen gold zeker dat "*Gutta cavat lapidem, non vi sed saepe cadendo*" (de druppel holt de steen niet uit door zijn kracht, maar door zijn gestadig vallen, Ovidius). Maar dan mag niemand de kraan dicht draaien. Bij de wettelijk aangestelde inspecteurs was dit alleen door het hoogste gezag mogelijk, en niet door de directies van de door hen geëvalueerde inrichtingen. Zij bekleedden een beschermde lijnfunctie.

De laatste onderzoeksvraag: *"Wat zijn de consequenties van een programma-evaluatiesysteem in de praktijk en heeft de ingewonnen informatie geleid tot een verbetering van het functioneren"*, kan nu aldus beantwoord worden: *"De instelling was bereid een dergelijk systeem in te voeren en (gedeeltelijk) de lasten ervan te dragen, maar de ingewonnen informatie en feedback heeft niet tot beleidsveranderingen geleid, voorzover dit door de onderzoeker kan worden beoordeeld."* Hoewel deze conclusie in eerste instantie alleen voor het onderhavige project geldt, mag worden verwacht dat overeenkomstige problemen zich ook in andere instellingen zullen kunnen voordoen. Daarvoor wordt er door programma-evaluatoren te vaak gemeld dat hun gegevens niet worden benut. Bovenstaande verklaring in organisatie-ontwikkelingstermen kan althans een deel van deze controverse oplossen.

11.4 OVERIGE EVALUATIE THEMA'S.

Naast het manco van het niet aanwezig zijn van gespecificeerde doelstellingen (zie 11.2) en van een organisatie-ontwikkelingsmodel (zie 11.3) zijn nog enkele kritische kanttekeningen te plaatsen bij dit evaluatieproject, welke van belang kunnen zijn voor degenen die een dergelijke project eveneens zouden willen uitvoeren.

Gedurende de periode 1980-1984 liepen programma-evaluatieve en onderzoekstechnische doelstellingen parallel. Naast het extra tijdsbeslag dat dit betekende voor met name patiënten, administratief personeel en onderzoeker, leverde dit voor de organisatie zelf weinig extra belasting op. Toch ontstond er een houding van: Als het onderzoek achter de rug is, zullen we wel weer eens iets met de evaluatie-gegevens doen. De grote hoeveelheid verzamelde gegevens (waarvan slechts een klein gedeelte werd gerapporteerd) betekende toch een drempel (of alibi) om inhoudelijk naar de resultaten te kijken. Dit leidt tot een vermenging van doelstellingen en risico's van complicaties. Bij een combinatie van programma-evaluatie en andere onderzoeksvraagstellingen moeten de belangen goed worden gescheiden.

Wordt alleen een programma-evaluatie project uitgevoerd, dan moet het argument van spaarzaamheid (Van den Hout, 1984) en *de wet van de verminderde meeropbrengst* in het oog worden gehouden en een te uitgebreide gegevensverzameling worden vermeden.

Een vermenging van doelstellingen (evaluatie en onderzoek) heeft ook een vermenging van rollen tot gevolg: de onderzoeker is ook evaluator en de evaluatie-gegevens worden ook voor het eigen onderzoek gebruikt. Hierdoor wordt zijn betrokkenheid met het project zeer groot, het wordt "zijn" onderzoek, het is (dus) belangrijk en moet (dus) worden gehonoreerd. Een dergelijke binding kan nadelig werken, hetzij dat de druk op het systeem - en dus ook de weerstand - te groot wordt, hetzij dat het project niet wordt beëindigd, wanneer dit is geïndiceerd.

Uit het bovenstaande vloeit een ander risico voort. Zoals in hoofdstuk 3 is vermeld, bestaat de derde weerstandsreactie uit het *overvragen* van de onderzoeker. Men weet dat er nog veel meer gegevens zijn en allerlei mogelijke verbanden kunnen worden geopperd die met extra analyses kunnen worden nagegaan. In een half uur kunnen zo meer vragen worden gesteld dan in maanden zijn te beantwoorden. Deze vragen appelleren ook aan het eigen gevoel van de onderzoeker: hij zal graag deze informatie verschaffen. Tevens hoopt hij dat dan zijn resultaten dan worden gebruikt. En zo valt hij in de kuil die hij zelf heeft helpen graven. Hij zal *geen afdoende verklaring* kunnen geven voor het *waarom* van de geconstateerde situatie en gaat dit dan ook als eigen falen zien. Op nadere vragen ingaan, zal zelden betekenen dat de gegevens wel worden gebruikt. Voor de reactie van overvragen moet derhalve gewaakt worden.

Zoals in 11.1 is gesteld, staat programma-evaluatie nog in de kinderschoenen, en in Nederland zelfs nog in de box; op zeer beperkte schaal wordt programma-evaluatie verricht en een verruiming van de mogelijkheden is dringend gewenst. Er dient evenwel gewerkt te worden aan een bundeling van de opgedane ervaringen en van die welke nog zullen volgen. Programma-evaluatoren hebben de ondankbare taak om juist die zaken aan te tonen die in een instelling niet goed functioneren. Dit zal altijd spanning veroorzaken, moet dit ook doen om veranderingen te kunnen genereren. In 11.3 is betoogd dat organisatie-psychologische deskundigheid is vereist, om de resultaten te kunnen (laten) vertalen in beleidsveranderingen. Publicatie van de ervaringen van dit proces werd ook reeds door Beenackers (1984) betoogd, ter lering, om zo de kinderziekten te kunnen overwinnen. Vandaar ook dit hoofdstuk.

Deze ervaringen strekken zich ook uit tot de formele en informele aspecten van de interactie tussen evaluator en geëvalueerden. Ten aanzien van de *formele zaken* is het wenselijk bij de start van het project goede (schriftelijke) afspraken te maken over de rol en positie van de evaluator, de overlegstructuur e.d. Ook over de *feitelijke producten* van het project: de ontwikkelde instrumenten, de (vertrouwelijk) ingevulde vragenlijsten en de gecomputeriseerde gegevensbestanden moet van meet af aan duidelijk zijn bij wie het (auteursrechtelijk) eigendom berust, en in hoeverre er door anderen van de gegevens kan worden gebruik gemaakt. Achteraf zal dit vaak moeilijk te regelen zijn. Bij de onderzoeker rust ook een ethische verantwoordelijkheid om de belangen van de patiënten te beschermen met betrekking tot de vertrouwelijk verworven informatie.

De formele relatie-aspecten die zullen worden beïnvloed door de controlerende taak van de evaluator, moeten bespreekbaar zijn met de medewerkers van de geëvalueerde instelling, zeker wanneer zich hierin wrijvingen voordoen. Duidelijk moet zijn dat de taak van de evaluator een bepaalde functionele houding tot gevolg heeft, die zich niet persoonlijk tegen bepaalde medewerkers of het instituut als zodanig richt. Een dergelijke houding vertoont veel overeenkomst met die van een psychotherapeut ten opzichte van zijn patiënt: distantieënd en objectiverend, maar tegelijkertijd met een grote betrokkenheid. Deze twee benaderingen zijn soms tegenstrijdig, en voor de patiënt verwarrend, terwijl ook de therapeut steeds weer het juiste evenwicht moet zien te vinden.

In een groep van collega's zouden ook deze aspecten van het functioneren van evaluatoren aan een vorm van intercollegiale toetsing onderworpen moeten worden. De (slapende) Werkgroep Evaluatie Geestelijke Gezondheidszorg (WEGG, NcGv Utrecht) zou hiervoor een goed forum kunnen zijn. Een bundeling van ervaringen in een dergelijke werkgroep is in Nederland dringend gewenst, en deze zou geen slapend bestaan mogen leiden.

Hiermee zijn de belangrijkste ervaringen van dit project geëvalueerd. Naast de specifieke factoren van de geëvalueerde instelling zijn er ook algemene kenmerken aan dit project te onderscheiden, die voor het beoefenen van programma-evaluatie als zodanig van belang zijn. De ervaringen van dit project kunnen derhalve van algemener nut zijn.

Tenslotte, zou uit deze evaluatie van P.O.P.E. ten onrechte de conclusie getrokken kunnen worden, dat programma-evaluatie een louter frustrerende aangelegenheid is voor alle betrokkenen. Deze indruk is dan onjuist. De coöperatie van hoegenaamd alle psychotherapeuten was zeer goed, alsmede de sociale contacten met hen.

Ook de medewerking van de patiënten, de belangrijkste informanten, was uitstekend, zoals reeds bleek uit de antwoordpercentages (zie hoofdstuk 8). Voor velen van hen had de evaluator ook de functie van quasi-Patiënten-Vertrouwens Persoon, een buitenstaander die erbij hoorde, waar ook kritiek bij geuit kon worden. Soms nog lang na ontslag namen ze met hem contact op om te laten horen hoe het ging of om advies te vragen. Een dergelijke functionaris lijkt in een behoefte te voorzien.

De derde participant, de evaluator zelf, kijkt pro saldo eveneens tevreden terug op het evaluatieproject. In 1977 werd de inleiding van het eerste interne verslag (Sijben, 1977a) besloten met: *"Of al deze inspanningen inderdaad effect zullen sorteren, weten we niet; in ieder geval geldt voor de auteur: Et conari pulchrum est"* (p. 1). Deze inspanningen hebben zekere effecten gesorteerd, naast dit verslag. De opgedane ervaringen leiden tot aanbevelingen voor andere programma-evaluatieprojecten (zie 11.5). En inderdaad, zoals boven dit hoofdstuk reeds staat vermeld: het pogen was ook schoon.

11.5 AANBEVELINGEN.

Een periode van ruim 8 jaar overziend, waarin in een praktijksetting ervaring is opgedaan met het evalueren van het resultaat van de behandeling, o.a. ook middels vergelijkende meetinstrumenten, zullen de belangrijkste gegevens en ervaringen hier in een aantal aanbevelingen worden vervat.

1. De informatie-inhoud.

Om de effectiviteit van de behandeling te beoordelen, is de zgn. "Target complaints"-methode een geschikt instrument. Daarnaast dient informatie verzameld te worden over satisfactie en dienen de diverse procedures rond opname en ontslag geëvalueerd te worden. Gelet op de grote invloed van de behandelaar op het resultaat dient zijn functioneren veel evaluatieve aandacht te krijgen. Het is evenwel noodzakelijk, dat patiënten vrij deze informatie kunnen geven. Een vertrouwelijke ontslagvragenlijst is hiervoor een mini-

male conditie. Naast deze meer algemeen geldende informatie, dient binnen de instelling gespecificeerd te worden welke doelen geëvalueerd moeten worden en aan welke feedback verder behoefte bestaat. Diagnosen zijn een weinig bruikbare (evaluatie)meting.

2. De informanten.

De belangrijkste informatieverstrekkers zullen de patiënten zelf zijn. Uit dit en ander onderzoek is gebleken, dat behandelaars te onbetrouwbaar (in wetenschappelijke zin) zijn om hun eigen activiteiten en het functioneren van de patiënt te kunnen beoordelen. Hun gegevens zijn hoogstens zijdelings bruikbaar.

3. De meetmomenten.

De belangrijkste meetmomenten liggen uiteraard aan het begin en aan het einde van een behandeling. Wil de evaluatie een formatief karakter hebben, dan zal ook tenminste één maal tijdens de behandeling expliciet door de patiënt geëvalueerd moeten worden. Na beëindiging zijn twee follow-up metingen een minimale vereiste: na drie maanden omdat deze periode het grootste verval te zien geeft en er ev. nog sturend kan worden opgetreden ten aanzien van bijv. het beter regelen van nazorg en na een jaar of meer. Bij een zeer lange follow-up periode wordt het risico van een te lage respons groter. Een grote betrokkenheid van de evaluator bij de patiënten (persoonlijk contact en uitleg) doen de responsratio's zeker stijgen.

4. De feedback.

Al deze inspanningen zijn alleen zinvol indien er werkelijke behoefte is aan feedback. Deze behoefte kan vooraf gepeild worden, bijv. met behulp van de "AVICTORY"-vragenlijst (Kiresuk, 1976 Larsen z.j.). Aan een Nederlandse vertaling wordt momenteel gewerkt. Uit het antwoordprofiel kan blijken hoezeer en met welke motieven men bereid is programma-evaluatie te accepteren. Het risico van sociaal wenselijke invulling is echter groot. Toch kan zo de discussie over behoefte aan feedback en mogelijke weerstanden worden begonnen.

Ook tussentijdse evaluatie dient uit te monden in wijzigingen van de behandeling indien dit noodzakelijk blijkt. Hier geldt met name: beter ten halve gekeerd dan ten hele gedwaald. Binnen de instelling zal een attitude aanwezig moeten zijn die op veranderingen gericht is, zowel op management-niveau als bij de medewerkers, wil de feedback gebruikt kunnen worden.

Zo nodig dient er organisatie-psychologische kennis aanwezig te zijn c.q. van buiten te worden betrokken, wanneer blijkt dat er geen attitude tot verandering aanwezig is. Het motiveren van alle betrokkenen en het aanbrengen van een dergelijke attitude dient expliciete aandacht te krijgen.

5. Intern of extern?

Programma-evaluatie dient een onlosmakelijk onderdeel van een organisatie te zijn, hetgeen per definitie een vorm van interne feedback betekent. Aangezien echter met name non-profit organisaties in grote mate verantwoording schuldig zijn aan de financier c.q. de maatschappij, moeten de effectiviteitsgegevens openbaar zijn. De evaluatieprocedure dient evenwel zodanig te zijn, dat interne weerstand gereduceerd wordt en dat de informatie toch betrouwbaar is. Tussen deze Scylla en Charibdis doorzeilen is geen gemakkelijke opgave. Een behoefte aan interne informatie teneinde het resultaat te verbeteren, is primair noodzakelijk om feedback zinvol te maken. Externe druk hierop middels een koppeling van inkomen (budget) en resultaat moet niet zonder meer worden afgewezen. Als het patiëntenbelang niet voldoende intrinsieke motivatie betekent om het behandelingsresultaat te verbeteren, moet overwogen worden hoe dit met extrinsieke middelen kan worden versterkt. Deze druk moet niet gelegd worden op het middel (de evaluatie) maar op het doel (de prestatie).

6. De positie van de evaluator.

De evaluator bekleedt een functie die het midden houdt tussen die van een hulpverlener en een belastinginspecteur. Hij dient met argusogen te speuren naar onjuistheden, maar met de bedoeling dat deze worden verbeterd. Is hij teveel inspecteur, dan roept dit verzet op en proberen de informanten gegevens verborgen te houden. Is hij te zeer een hulpverlener, dan let hij teveel op de positieve aspecten van het functioneren en kunnen hem die zaken ontgaan die verandering behoeven. Voor intercollegiale toetsing alsmede voor overleg over de andere problemen die programma-evaluatoren in hun werk tegenkomen, dient een landelijk forum te bestaan.

Organisatorisch dient zijn functie dusdanig te zijn dat de continuering van zijn werkzaamheden is gewaarborgd en zijn boodschap wordt gehonoreerd. Anders wordt niet het programma maar de evaluator gewogen en te licht bevonden.

7. Implementatie van programma-evaluatie.

Er blijkt een toenemende behoefte aan programma-evaluatieprojecten, zowel in de geestelijke gezondheidszorg-instellingen als op ministerieel niveau. Inmiddels is in Nederland enige (divergente) expertise aanwezig ten aanzien van opzet en uitvoering van programma-evaluatieprojecten. Deze ervaring dient gebundeld te worden (zie ook de vorige aanbeveling). Voor een goede dissimiatie van de programma-evaluatiekennis, meetprocedures en ervaringsgegevens zou een landelijke coördinatie van belang zijn. Instellingen met programma-evaluatieve vragen zouden hier terecht moeten kunnen en ook daadwerkelijke hulp moeten kunnen krijgen, zowel voor personele als voor materiële (automatiseringsapparatuur, e.d.) ondersteuning. Zo zal programma-evaluatie

wellicht een evaluatiesysteem worden, dat in de geestelijke gezondheidszorg de taak zal overnemen die de inspecteurs in de vorige eeuw hadden, althans in evaluatief opzicht.

11.6 TENSLOTTE.

Wordt uiteindelijk teruggekeerd naar het schijndilemma tussen onderzoek en programma-evaluatie (hoofdstuk 2), dan biedt programma-evaluatie ook in dat opzicht perspectieven. De kloof tussen onderzoek en praktijk zou middels programma-evaluatie kunnen worden overbrugd vanuit de praktijkoever. Aldus kan betere informatie verkregen worden over de effectiviteit van bepaalde behandelingen, eventueel zelfs over causale relaties tussen doel en middel, dan experimenteel onderzoek dit in de loop der jaren heeft kunnen doen. Programma-evaluatie waarborgt een continue stroom van input- en output/effect gegevens. Met een goede registratie van behandelingsvariabelen kan vervolgens de specifieke effectiviteit van een bepaalde interventie worden nagegaan bij een steeds groeiend aantal patiënten.

Programma-evaluatie wordt dan een platform waarop onderzoek naar bijv. bepaalde behandelingsstrategieën kan worden gebouwd. Een dergelijke constructie wordt binnen de afdeling Psychiatrie van het St. Radboudziekenhuis ingevoerd. Van hieruit kunnen deze evaluatieprocedures alsmede worden verspreid naar de geaffiliëerde ziekenhuizen en andere geïnteresseerden. En aldus bereikt dit project uiteindelijk zijn doelstelling, en zelfs meer dan dat. Een constatering die dit verslag met voldoening kan besluiten.

Samenvatting

Doe wel en zie niet om, is voor de geestelijke gezondheidszorg een achterhaald spreekwoord. In toenemende mate, mede onder druk van bezuiniging en de noodzaak tot het afleggen van verantwoording, bestaat er een extrinsiek motief om aan te tonen in hoeverre er wordt *"wel gedaan"*. Er moet worden *"omgezien"* naar het *"weldoen"*. Deze tendens is medio 1960 reeds in de U.S.A. duidelijk aanwezig, en leidt tot een sterk groeiende belangstelling voor een nieuwe of vernieuwde vorm van evaluatie-onderzoek: *"Program Evaluation"*.

Eind jaren zeventig raakt deze vorm van onderzoek ook in Nederland geleidelijk in zwang onder de naam *"Programma-Evaluatie"*. Deze vorm van toegepast sociaal onderzoek wordt gedefiniëerd als: *het vaststellen van de waarde van een programma door de effecten ervan te vergelijken met de doelen die hieraan vooraf zijn gesteld*. In deze dissertatie wordt verslag gedaan van een programma-evaluatieproject dat gedurende ruim 8 jaar in een kliniek voor gedragstherapie is uitgevoerd. Aan de beschrijving van dit project en de resultaten en conclusies ervan gaat een uitgebreide theoretische studie vooraf.

Deel A: Programma-evaluatie in theorie.

Hoofdstuk 1. Omzien naar weldoen.

Staatstoezicht als programma-evaluatie

Programma-evaluatie lijkt een product van de groeiende behoefte aan *verantwoording* van de besteding van overheids-gelden. Een katalysator hierbij is dan de mogelijkheid van geautomatiseerde gegevensopslag en -bewerking. Toch blijkt dit niet geheel het geval. In dit hoofdstuk is nagegaan of en hoe in het verleden in Nederland de overheid heeft omgezien naar het weldoen in de geestelijke gezondheidszorg. Dit *historisch onderzoek* voert terug tot de verzelfstandiging van Nederland. Onder het koningsschap van Koning Willem I werd een begin gemaakt met het verbeteren van de zorg voor de krankzinnigen. Zij waren dikwijls ongedifferentieerd opgesloten in de oude leprahuizen, samen met anderen: daklozen, werkelozen e.d. Hun huisvesting en verzorging was erbarmelijk. Gedurende de periode 1800-1841 wordt voornamelijk de situatie in kaart gebracht middels opeenvolgende enquêtes en bezoeken van de Minister van Binnenlandse Zaken samen met referendaris Feith. Eind jaren dertig worden concrete voorbereidingen getroffen voor wetgeving, welke in 1841 resulteert in de eerste zgn. *Krankzinnigen Wet*. Hiermee trachtte men de huisvesting en behandeling te regelen. Deze wet kan worden opgevat als een doelstelling. De daarin voorgeschreven situatie stond nog ver af van de geldende praktijk. Teneinde

te kunnen nagaan in hoeverre de reële situatie met de wettelijke bepalingen overeenstemde, werden in 1842 de eerste twee *inspecteurs van het staatstoezicht op de krankzinnigen* benoemd, Prof.Dr. J.L.C. Schröder van der Kolk en referendaris C.J. Feith. Zij kunnen worden beschouwd als de eerste twee *programma-evaluatoren* in Nederland. Hun eerste inspectietocht toonde aan, dat slechts drie van de 28 inrichtingen aan de vereisten van de nieuwe wet voldeden. Zij rapporteerden dienovereenkomstig aan de minister, in de hoop en verwachting dat deze iets aan deze situatie zou doen. Maar 's lands financiën lieten een drastische herziening niet toe.

In de opeenvolgende jaren blijven de inspecteurs hun inspecties uitvoeren - de oorspronkelijke aantekeningen zijn nog bewaard - en in jaarverslagen rapporteren. Geleidelijk blijkt deze evaluatie en feedback effect te sorteren. De eenvoudige effectmaat: het *verbeteringspercentage* begint geleidelijk te stijgen en steeds meer inrichtingen gaan aan de voorwaarden van de wet voldoen. Het aantal bewaarplaatsen (voor zgn. onbehandelbare, chronische patiënten) daalt en in 1886 wordt de laatste gesloten: het Cellenbroederhuis in Nijmegen.

Een belangrijk verschijnsel is de exponentiële groei van het aantal opgenomen patiënten: van ca. 1000 in 1845 naar bijna 5000 in 1885. Dit had uiteraard gevolgen voor de evaluatieve functie van de inspecteurs, wier aantal niet evenredig groter werd. Van directe inspectie op *patiëtniveau* moest hun toezicht *afstandelijker* en meer op *instituutsniveau* plaatsvinden. Met behulp van de evaluatieve maat: de *herstelratio* (aantal hersteld ontslagen/aantal opgenomen patiënten) wordt nagegaan of de behandeling voldoende resultaat gaf. Ook werden vergelijkingen tussen instellingen gemaakt.

In 1884 wordt de Wet zelf herzien. Sommige doelstellingen waren achterhaald of verdienden bijstelling. De inspecties gaan door, en hun jaarverslagen groeien tot omvangrijke boekwerken. Ook het aantal patiënten blijft gestaag groeien: van ruim 5000 in 1886 tot bijna 26000 in 1936. Echter, sinds 1900 begint de herstelratio te dalen, en deze verdwijnt in 1909 uit de jaarverslagen. Bij een registerwijziging in 1930 vervalt de aanduiding van de mate van herstel bij ontslag. De evaluatieve maat wordt afgeschaft, wanneer deze juist het meest nodig blijkt te zijn. Ook dit is een herkenbaar fenomeen bij programma-evaluatie: de bewaker gaat naar huis als er wordt ingebroken. In 1936 stopt de stroom jaarverslagen.

In 1975 verschijnt weer het eerste verslag over de periode 1969-1974. Met behulp van het *Patiënten Intramurale Geestelijke Gezondheidszorg (P.I.G.G.)* registratiesysteem, dat in 1967 is opgezet, worden ook weer landelijke gegevens verzameld. Het aantal psychiatrische patiënten in de Algemene Psychiatrische ziekenhuizen blijkt op het niveau van 1936 te staan: 26000 en gaat in de jaren daarna dalen tot ruim 21000 in 1983. De aard van de verslagen is geheel anders dan die van de collega's uit de vorige eeuw. Naast de klinische zorg heeft de inspectie nu ook het toezicht op alle andere, inmid-

dels nieuw ontstane residentiële, semi-murale en ambulante voorzieningen in de geestelijke gezondheidszorg. Het staats-toezicht is meer gericht op de "*inspannings*"-zijde van het functioneren: personele, materiele voorzieningen, dan op het *effectgedeelte*: de resultaten van de behandeling. Geconstateerd wordt o.a. uit de P.I.G.G.-cijfers, dat het aantal verblijfspatiënten in de algemeen psychiatrische ziekenhuizen toeneemt en er gaan anno 1983 weer stemmen op om zgn. *bewaarplaatsen nieuwe stijl* voor deze patiënten op te zetten. Ook in ander opzicht lijkt de geschiedenis zich te herhalen, wanneer weer gevallen van langdurige separatie (isolering) worden geconstateerd.

Uit de geschiedenis van het Staatstoezicht op de krankzinnigen zijn in programma-evaluatief opzicht diverse lessen te leren:

- Er dienen toetsbare (behandelings-/effect-) *doelstellingen* geformuleerd te worden.
- Deze moeten in een hanteerbare maat worden *geoperationaliseerd*.
- Er moeten waarborgen zijn dat deze evaluatieve informatie als *feedback* wordt gebruikt zodat het behandelingsresultaat *verbetert*.
- Alle participanten: overheid, geëvalueerde en evaluerende instelling moeten dusdanig *samenwerken* dat het evaluatieproces zelf verbetert.

Een (landelijk) programma-evaluatiesysteem zal het "*omzien naar weldoen*" gestalte kunnen geven, hetgeen het welzijn van de zwakste groep, de patiënten, ten goede zal komen.

Hoofdstuk 2. Is evalueren ook onderzoeken?

In hoofdstuk 2 wordt de (academische) vraag beantwoord of programma-evaluatie tot het terrein van het sociaal wetenschappelijk onderzoek kan worden gerekend. Hiertoe wordt uitgegaan van de definitie van onderzoek en zijn verschijningsvormen zoals deze door De Groot (1961) zijn beschreven. Het door hem gesignaleerde "*evaluatieprobleem*", waaronder het evalueren van de effecten van een behandeling wordt begrepen, rekent hij tot het basismodel van het toetsend onderzoek. Ten onrechte is hieronder steeds experimentele toetsing van hypothesen begrepen.

Ook Hofstee (1980), die onderzoek omschrijft volgens het zgn. *weddenshapsmodel*: nagegaan moet worden wie *gelijk* heeft, kent aan het evalueren van effecten een belangrijke plaats toe. Hij waarschuwt echter voor toegepast onderzoek, omdat dit meer voor "*public relations*"-doeleinden wordt gebruikt en als onderzoeksinvestering maatschappelijk irrelevant is, omdat de onderzochte partij zijn ongelijk niet zal wensen te erkennen.

Het verschil tussen *onderzoek* en *evaluatie* is vooral gelegen in de aard van de informatie die verkregen wordt en het

doel van deze wetenschappelijke activiteit. Onderzoek is meer *nomothetisch*, het gaat om *kennis* te verwerven en wetmatigheden te ontdekken; evaluatie is vooral *ideografisch* en is bedoeld om mogelijk '*kunnen*' aan te tonen. Er wordt een onderscheid gemaakt tussen *behandelingsonderzoek* en *behandelingsevaluatie*. Hierbij verwisselt het primaire en secundaire doel (wetenschappelijke kennis vermeerderen versus verbetering van behandelingseffecten). Dit heeft ook methodologische consequenties: een omkering van de *afhankelijke* en de *onafhankelijke variabele*. Dit heeft vervolgens consequenties voor de toetsingsprocedure (normatief of absoluut) en de interpretatie van de resultaten (statistisch significant of klinisch relevant).

Behandelingsevaluatie kan zich in meerdere verschijningsvormen voordoen: *summatief* of *formatief* en "*process*" of "*outcome*"-evaluatie. Programma-evaluatie is een vorm van behandelingsevaluatie. Er wordt hoofdzakelijk "*outcome*" of effect-informatie verzameld, en de gegevens worden voor formatieve doeleinden gebruikt, zeker op instellingsniveau. Geconcludeerd wordt, dat de discussie tussen sociale wetenschappers of programma-evaluatie al dan niet wetenschappelijk onderzoek is, als onterecht en zinloos kan worden bestempeld.

Hoofdstuk 3. Programma-evaluatie, wat is dat?

Programma-evaluatie kan gedefiniëerd worden als: *het vaststellen van de waarde van een programma door de effecten ervan te vergelijken met de doelen die voor dit programma zijn gesteld*. Een programma is iedere activiteit die erop gericht is om sociaal, economisch of psychisch welzijn te verbeteren en zal over het algemeen uit overheidsmiddelen worden gefinancierd. Het *doel* van programma-evaluatie is het *geven van feedback*, zodat het functioneren van het programma verbetert.

Programma-evaluatie kan zich op verschillende onderdelen van een programma richten: de *inspanning* (budget, personeel, doelgroep, etc.), de *prestatie* (effecten, satisfactie, e.d.), de *adequaatheid* (relatie tussen effect en behoefte), *doelmaticheid* (relatie tussen inspanning en prestatie) en *proces* (waarom zijn deze resultaten bereikt). Deze laatste wordt doorgaans niet tot het domein van de programma-evaluatie gerekend.

De evaluator kan of zelfs moet zich (in eerste instantie) *beperken* tot het geven van *feedback*, het beschrijven van de status quo in relatie tot de doelstellingen. Over het *waarom* ervan zal hij hoogstens zijdelings informatie hebben, en altijd te weinig om deze waarom-vraag (afdoende) te kunnen beantwoorden. Gelet op de belangentegenstellingen tussen de geestelijke gezondheidszorginstellingen enerzijds en de overheid/financier anderzijds wordt een onderscheid gemaakt tussen *interne*, door de instelling zelf uitgevoerde programma-evaluatie en *externe* programma-evaluatie, welke

door (en voor) een externe instantie wordt verricht. Een andere indeling is aangebracht in het niveau waarop programma-evaluatie zich richt: patiënt, programma, instelling of groep van instellingen (regionaal of landelijk). Ten aanzien van de procedure van programma-evaluatie is de aard (intern/extern) of het niveau (patiënt tot landelijk) niet van grote betekenis.

De evaluatie-procedure van programma-evaluatie doorloopt een *cyclisch proces*:

- In een *beleidsplan* moeten de doelen van het programma worden gespecificeerd alsmede de concrete c.q. objectieve informatiebehoefte per beleidsniveau.
- Vervolgens wordt in een *evaluatieplan* deze informatiebehoefte geoperationaliseerd in meetinstrumenten en -procedures: wie zal welke informatie waarover verschaffen.
- Hierna wordt dit plan geleidelijk *geïmplementeerd*. Bij voorkeur volgens een zgn. *ui-model*; d.w.z. dat van buiten naar binnen wordt gewerkt. Eerst worden de intake/opname en de beëindiging/ontslag (met bij voorkeur een of meer follow-up metingen) geëvalueerd. Daarna kunnen ook interne, tussentijdse evaluaties ingevoerd en behandelingsgegevens verzameld worden.
- *Evaluatie en feedback* is het laatste stadium van de eerste cyclus. De verzamelde informatie wordt geanalyseerd, gerelateerd aan de doelen en gerapporteerd aan de belanghebbenden. Deze kunnen, indien noodzakelijk, het beleidsplan veranderen en een nieuwe *evaluatiecyclus* kan starten.

Deze feedbackboodschap zal niet altijd met gejuich worden ontvangen. Programma-evaluatoren klagen over het feit, dat met hun gegevens te weinig wordt gedaan. Diverse *weerstand-reacties* worden door hen opgeroepen. Drie bekende zijn: *wegverklaren*, *ontkennen* en *overvragen*. Programma-evaluatie is een spiegel die een instelling wordt voorgehouden. Het is de taak van het eigen management eventuele fouten die wordenesignaleerd, te (laten) behandelen.

Hoofdstuk 4. Het meten van psychotherapie effecten.

Een van de moeilijkste problemen bij het doen van onderzoek op het gebied van de psychotherapie is het *operationaliseren* van het resultaat. In de loop der jaren is een grote diversiteit van meetinstrumenten ontwikkeld en gebruikt om in doorgaans (experimenteel) vergelijkend onderzoek het effect van de onderscheiden behandelingen te meten. Deze metingen variëren van projectieve technieken tot vragenlijsten, van gedragsobservaties tot globale beoordelingen van het resultaat. Ze kunnen worden gewaardeerd op dimensies als reactiviteit en specificiteit naast de gebruikelijke psychometrische parameters. Gelet op de aanbevelingen die volgden uit enkele grote meta-evaluatieve studies, worden drie evaluatiemethoden uitgebreid beschreven: *doelevaluaties*, *psychologische vragenlijsten* en *subjectieve effect/satisfactie-beoordelingen*.

Van de geïndividualiseerde doelevaluatiemethoden zijn twee procedures op grotere schaal bekend geworden: "*Target Complaints*" en "*Goal Attainment Scaling*". Vooral de eerstgenoemde lijkt een methode die voor effectonderzoek in de praktijk goed bruikbaar is. Deze is weinig arbeidsintensief en kost ook weinig (professionele) begeleiding. Voor een permanente toepassing van de G.A.S. gelden deze twee kenmerken niet. Deze methode verdient constante aandacht, hulp en begeleiding en kost de invuller veel tijd. Daarom is G.A.S. minder aangewezen als evaluatie-procedure in de reguliere praktijk. Een vereenvoudigde versie, de *Individuele Doel Evaluatie (Meting)* biedt wellicht meer mogelijkheden.

De *psychologische vragenlijsten* zijn het frequentst als evaluatie-instrument gebruikt. Bij nadere beschouwing blijken de in Nederland meest gebruikte lijsten nauwelijks te voldoen aan de normen van betrouwbaarheid en validiteit, terwijl zij juist om deze eigenschappen - naast het gebruiksgemak - worden verkozen. Het voordeel is wel dat deze gestandaardiseerde instrumenten vergelijkbare informatie verschaffen.

Subjectieve beoordelingen van het effect en de tevredenheid over de behandeling zijn wellicht het meest in de praktijk gebruikt. Met behulp van zelf ontworpen vragenlijsten wordt aan de patiënt gevraagd het *effect* te beoordelen en aan te geven hoe *tevreden* hij is over het gebodene. Door de grote variëteit aan deze zgn. "home made" lijsten, leveren deze nauwelijks vergelijkbare informatie op.

Een instrument dat in tal van onderzoek is uitgetoetst en via analyses is verijnd, is de *Client Satisfaction Questionnaire (C.S.Q.)*. Deze is ook in het Nederlands vertaald. Deze lijst lijkt goed bruikbaar om hiermee, naast andere gegevens, het resultaat van een behandeling te evalueren.

Evaluatie van psychotherapie-effecten behoeft bij voorkeur een "*multitrait-multimethod*" benadering. Psychotherapie kan tot uiteenlopende effecten leiden en naast het *directe behandelingsresultaat* (doelmetingen) is ook subjectieve beoordeling van de *tevredenheid* van de patiënt waardevolle informatie om in een programma-evaluatie te verzamelen. Daarnaast kan dit op de lange duur vergelijkende informatie opleveren tussen de verschillende behandelingsmethoden en kan aldus een bijdragen leveren aan het onderzoek naar (causale) verbanden tussen doel en middel.

Deel B. Programma-evaluatie in praktijk.

Hoofdstuk 5. Het werkterrein, Kliniek Overwaal.

Gedurende een periode van ruim 8 jaar is in een kliniek voor psychotherapie (30 bedden) een programma-evaluatie-project uitgevoerd. In dit hoofdstuk worden de belangrijkste kenmerken van deze kliniek beschreven. Deze zijn:

- De behandelingsstrategie is voornamelijk *gedragstherapeutisch* georiënteerd.

- De behandeling is sterk *geïndividualiseerd*. Een psychotherapeut voert, in samenwerking met een sociotherapeut, een individueel opgesteld behandelingsplan uit. Daarnaast wordt van groepstherapieën gebruik gemaakt, zoals gespreksgroepen, bewegingstherapie, creatieve therapie.
- De behandeling is qua tijd en personele bezetting zeer *intensief*: per 4 à 5 patiënten is een psychotherapeut en sociotherapeut (B-verpleegkundige) beschikbaar. Daarnaast kan nog een beroep worden gedaan op bewegings- en creatief therapeuten.
- De behandelingsduur dient *kort* te zijn en zoveel mogelijk aan de zgn. Al-norm te voldoen (minder dan 6 maanden).
- De behandeling is geïndiceerd voor *neurotische volwassen* patiënten die niet of niet meer met ambulante hulpverlening geholpen kunnen worden.
- Na deze intensieve klinische zorg volgt in principe (weer) *ambulante hulpverlening*. Een behandelingsdoel is, dat de patiënt hier weer met vrucht gebruik van kan maken.

Hoofdstuk 6. Project Overwaal Programma-Evaluatie.

In 1976 werd op Overwaal op eigen verzoek een klinisch/researchpsycholoog gedetacheerd door de afdeling Psychiatrie van de Katholieke Universiteit Nijmegen met de bedoeling een *evaluatiesysteem* op te zetten dat zou dienen voor *interne feedback* van het resultaat van de behandeling.

Het door hem uitgevoerde Project Overwaal Programma-Evaluatie is in een viertal chronologische fasen te onderscheiden:

- Een *retrospectief follow-up onderzoek* over de periode 1974-1976. Hierin werd ervaring opgedaan over de te gebruiken ontslag- en follow-up-vragenlijst.
- In de periode 1976-1980 werd een *permanent evaluatie-onderzoek* verricht waarbij bij opname, ontslag en 3 follow-up-momenten patiënten en behandelaars informatie verschaften over tal van aspecten van het verloop en resultaat van de behandeling.
- Van 1980 tot 1983 werd een *vergelijkend onderzoek* uitgevoerd tussen verschillende meetinstrumenten: - *doelevaluaties* door de patiënt, psychotherapeut en externe beoordeelaar, - *standaard psychologische vragenlijsten*, - *globale beoordeling van effect en satisfactie*. Deze evaluatieperiode werd tevens mogelijk gemaakt door een subsidie van het *Praeventiefonds* voor ondersteunend personeel.
- In 1983 werden de te verzamelen gegevens *sterk beperkt*: alleen de doelevaluaties van de patiënt, enkele psychologische vragenlijsten en de globale effect/satisfactie beoordelingen werden gecontinueerd.

Eind 1984 werd de detachering van de evaluator beëindigd, hetgeen als een gedeeltelijk einde van het project kan worden gezien. Na een administratieve continuering in 1985 is in 1986

de supervisie van het project overgenomen door een psychotherapeut van Overwaal. Door hem werd de verzamelde informatie sterk ingekort en gewijzigd. Dit kan als een feitelijke beëindiging van P.O.P.E. beschouwd worden. In dit onderzoeksverslag worden de gegevens over de periode 1974 tot medio 1985 verwerkt.

Hoofdstuk 7. De kwantitatieve gegevens geregistreerd.

Een evaluatieproject verloopt van *registratie* naar *evaluatie*. Eerst zal een informatiesysteem worden geïnstalleerd, zodat een goede beschrijving van de meer objectieve patiëntengegevens wordt verkregen. In hoofdstuk 7 staan deze gegevens van ruim 800 patiënten vermeld, in perioden van twee of drie jaren samengenomen om eventuele trendmatige veranderingen te kunnen signaleren.

Wordt de totale groep van patiënten in grote lijnen beschreven, dan ontstaat het volgende beeld:

- per jaar worden ca. 80 patiënten opgenomen, hun gemiddelde *verblijfsduur* is 140 dagen, met een standaarddeviatie van 84 dagen. Ca. 20% van de patiënten wordt reeds binnen 2 maanden ontslagen, hetgeen dikwijls als "drop out" kan worden aangemerkt. Dit percentage stijgt in de loop der jaren. Daarnaast verblijft ca. 25% langer dan 6 maanden op Overwaal, hetgeen volgens de zgn. Al-norm te lang is;
- gelet op de *demografische variabelen* kan de populatie van Overwaal-patiënten worden beschreven als: 60% vrouwen, 40% mannen, welke verhouding in de loop der jaren veranderde van 50-50 naar 67-33; de gemiddelde leeftijd is ruim 30 jaar, (sd. = 9); ruim 60% is gehuwd/samenwonend; ca. 55% behoort tot de laagste economische klassen en bijna een-vierde is huisvrouw.
Bijna 50% heeft als laatste een LO, LBO of MVO-opleiding gevolgd. De belangrijkste bron van inkomsten voor opname is een ziektewet/WAO-uitkering (37%), 25% was werkzaam en bijna 25% had geen eigen inkomen (huisvrouw/-man);
- de *aard* en de *duur* van de problemen worden vastgelegd in de diagnose bij opname en de mededeling van de patiënt op de vraag sinds wanneer hij zich van zijn problemen bewust is en sinds hoelang hij hiervoor reeds in behandeling is. Ca. 60% van de patiënten wordt gediagnostiseerd met *neurose*, met dwang en fobie als de meeste voorkomende. De klachten bestaan doorgaans reeds langer dan 3 jaar en hebben eveneens reeds vele jaren tot dysfunctioneren geleid. Meer dan de helft heeft ook reeds langer dan drie jaar hulpverlening, ruim een kwart zelfs meer dan 6 jaar. De belangrijkste verwijzers zijn de huisarts of het RIAGG. Wordt de hulpverlening vóór opname nader geanalyseerd, dan blijkt dat 40% reeds eerder opgenomen is geweest en 90% een of meermalen ambulant is behandeld. De totale periode sinds de start van een behandeling tot de opname op Overwaal is gemiddeld 42 maanden. Hiervan zijn echter 20 maanden zon-

der hulpverlening geweest, hetgeen een sterk onderbroken hulpverleningsgeschiedenis impliceert.

Deze registratiegegevens lenen zich enerzijds om na te gaan of de werkelijke patiëntengroep overeenstemt met de *doelgroep*, en anderzijds om naar relaties te zoeken tussen deze variabelen en effect-metingen, hetgeen in hoofdstuk 9 aan de orde komt.

Hoofdstuk 8. De evaluatie van de resultaten.

De resultaten van de drie opeenvolgende evaluatieperioden worden afzonderlijk beschreven, omdat de veranderingen in meetinstrumenten zoveel consequenties hadden, dat combinatie van gegevens niet mogelijk bleek.

In het *retrospectieve follow-up onderzoek (1974-1976)* werd van 125 patiënten informatie verkregen (responspercentage: 70%). Zij gaven aan dat de problemen bij ontslag bij 40% geheel/tamelijk goed verdwenen waren. Bij 45% was het resultaat gering, en bij 15% waren de problemen niet behandeld. Ondanks de niet zo hoge effectiviteit was de tevredenheid over de opname en behandeling betrekkelijk hoog. Er bleek een significant verschil in effect en satisfactie tussen patiënten die door de vaste staf versus door assistent-psychiaters (onder supervisie) waren behandeld. Door vraagartefacten en heterogeniteit van de onderzochte en te kleine groep zijn deze gegevens niet verder geïnterpreteerd.

Op basis van de ervaringen met deze, door de onderzoeker ontworpen, vragenlijst, werd met ingang van het najaar 1976 een *permanente evaluatie* uitgevoerd. De responsratio's van de uitgereikte c.q. verzonden lijsten waren hoog tot zeer hoog (71-100%). Deze ratio daalde met het verlengen van de follow-up periode.

De belangrijkste evaluatieve gegevens zijn:

- een *dalende effectiviteit* van de behandeling: in 1977 zegt bij ontslag 31% van de patiënten (N = 262) dat de problemen "een beetje" of niet verbeterd zijn of verslechterd. In 1980 is dit percentage gestegen tot 53%. Overeenkomstige vragen geven hetzelfde antwoordpatroon.
- De *satisfactie* over de behandeling is *hoog*.
- Ook de psychotherapeuten geven - een weliswaar kleinere - *daling van het behandelingsresultaat* aan.
- Bij de *eerste follow-up* blijkt een *groot verval* op te treden in de beoordeling van de status quo door de patiënt; zegt bij ontslag ca. 25% dat het "niet zo goed" of "slecht" gaat, bij de eerste follow-up is dit 40%. In het verdere verloop van de periode na ontslag daalt dit percentage weer naar 31%. Ook in de follow-up periode is het verminderde resultaat herkenbaar.
- De *hulpverleners na ontslag* geven *overeenkomstige oordelen* over de status quo van de patiënt. Het oordeel over het effect van de opname wordt in 1977 door 63% "zinnig" genoemd. In 1980 is dit nog 41%.

Samengevat blijkt uit de verzamelde gegevens een consistent beeld: er is een trendmatig dalend verloop van het behandelingsresultaat. Deze gegevens als zodanig zijn een product van een dergelijk descriptief evaluatieproject, en deze informatie kan een *feedbackfunctie* vervullen, het doel van een dergelijk evaluatie-project.

In 1980 worden de vragenlijsten drastisch herzien, het design blijft evenwel onveranderd. Een uitbreiding is met name de .us;geïndividualiseerde doel-evaluatie van de patiënt (Target-complaints-methode). De responsratio's blijven op hetzelfde niveau of zijn zelfs hoger dan in de vorige periode (72-100%).

- Het *behandelingsresultaat* zoals dit bij ontslag door diverse vragen wordt gemeten, blijkt wederom een dalende trend te vertonen, waarbij 1984 afwijkt in positieve zin. Over het geheel genomen geeft ca. 40% een gunstig oordeel, ca. 30% een matig oordeel en 30% een ongunstig oordeel (N = 320). Per gebruikte evaluatiemaat treden er verschillen op.

- De *satisfactie* is over het algemeen goed, evenals de beoordelingen van de psycho-/sociotherapeut.

- Het oordeel van de behandelaar is *gunstiger* dan dat van de patiënt. De prognose bij ontslag over hoe het met de patiënt na ontslag zal gaan is minder gunstig en daalt eveneens.

- In de *follow-up periode* doet zich hetzelfde verschijnsel voor als in de jaren 1976-1980: een sterke daling in de beoordeling van de status quo bij de eerste follow-up, welke in de loop van de tijd weer stijgt zonder evenwel het niveau van het oordeel bij ontslag te bereiken. Ook hier laat 1984 een gunstiger beeld zien dan de trend sinds 1981 zou doen verwachten.

- De oordelen van de *voortgezette hulpverleners* corresponderen met die van de patiënt, zij het dat deze eersten over het algemeen positiever zijn.

Worden de twee evaluatieperiodes: 1976-1980 en 1981-1985, vergeleken, dan blijkt dat kleine veranderingen in de formuleringen van de antwoordcategorieën van de vragen soms tot grote verschuivingen aanleiding zijn. De nieuwe lijsten geven een betere gradatie van de antwoordmogelijkheden.

Ondanks deze verschillen zijn, met enige voorzichtigheid, over de totale periode van ruim 8 jaar de resultaten als volgt samen te vatten:

- een stijgend percentage geeft een gering behandelingseffect aan: ca. 20% in 1977 tot 39% in 1985 zegt dat het opnameprobleem weinig/niet verbeterd of zelfs verslechterd is. Het gaat "niet zo goed" of "slecht" wordt door 19% in 1977 geantwoord, in 1985 is dit 31%;

- de resultaten bij follow-up geven een overeenkomstig beeld, waarbij de 1e follow-up het meest ongunstige effect laat zien: 40% zegt in 1977 dat het niet goed/slecht gaat, in 1983 is dit geleidelijk geklommen tot 52% om in 1984 weer te dalen tot 28%;

- de tevredenheid over de behandeling is constant zeer goed. Ca. 70-90% is (zeer) tevreden over diverse aspecten van de behandeling c.q. behandelaars;
- het regelen van nazorg is bij 80% van de patiënten drie weken vóór ontslag gebeurd, maar volgens de voortgezette hulpverleners is dit slechts bij 60% van de patiënten het geval. Laat-geregelde hulp leidt tot een ongewenst hiaat in de curatieve zorg, slechts bij ca. 30% sluit de voortgezette hulp direct aan op de klinische periode;
- in de follow-up periode blijken de problemen direct na ontslag sterk te verslechteren, waarna deze bij follow-up na 6 en 12 maanden weer geleidelijk verbeteren;
- een jaar na ontslag heeft 30% geen hulp meer nodig. De overige 70% heeft wel hulp of zou deze wensen.

Door het ontbreken van geëxpliciteerde doelen is de evaluatie van het programma "Overwaal" descriptiever dan deze zou moeten zijn. Desalniettemin wordt in dit hoofdstuk een beeld gegeven welke informatie een dergelijk project zou kunnen verschaffen.

Hoofdstuk 9. De resultaten geanalyseerd.

In P.O.P.E. zijn van ca. 800 patiënten een zeer grote hoeveelheid gegevens verzameld. Een dergelijk bestand van ca. 700 variabelen per patiënt leent zich ook voor nadere analyse. Deze concentreert zich vooral op het bestand sinds 1980. In dit hoofdstuk wordt eerst de *betrouwbaarheid* (interne consistentie en interbeoordelaarsbetrouwbaarheid) onderzocht. Beide zijn goed te noemen.

Vervolgens zijn de ontslag- en follow-up-vragenlijsten middels *factoranalyse* gereduceerd tot 23 factoren. De ladingen van de vragen op de factoren zijn steeds hoog (.50-.90), de verklaarde variantie per vragenlijst ca. 60%. Deze 23 factoren zijn als afhankelijke variabelen gebruikt om relaties met diverse onafhankelijke variabelen na te gaan.

Met behulp van *multivariate variantie-analyse* is getoetst of er systematische verschillen bestaan tussen: psychotherapeuten, diagnose bij opname en ontslagjaar. Van deze hoofdefecten geeft vooral de psychotherapeut een significant verschil te zien op de variabele therapie-effect en beoordeling psychotherapeut bij de ontslag- en follow-up-vragenlijst van de patiënt. De diagnose persoonlijkheidsstoornis geeft alleen bij de ontslagvragenlijst van de psychotherapeut een significant verschil op de variabele effect in vergelijking met de andere diagnosegroepen. Contaminatie mag hier niet worden uitgesloten, aangezien bij de beoordelingen van het effect door de patiënten met een dergelijke diagnose, geen verschil aantoonbaar is.

Tal van demografische variabelen van de patiënt zijn in *univariate variantie-analyses* onderzocht of deze met verschillen in effect verband houden. Dit blijkt enigermate het

geval met de variabele geslacht (mannen effectiever dan vrouwen) en het regelen/starten van voortgezette hulpverlening. Laat gestarte hulp na ontslag is gerelateerd aan weinig effect. Correlaties tussen duur van de klachten en hulpverlening voor opname blijken nauwelijks te bestaan.

Tenslotte is het behandelingsresultaat *criterium georieenteerd* geanalyseerd. Het effect is gedichotomiseerd: niet/wel effectieve behandelde patiënten. Van de totale groep patiënten blijkt 66% effectief behandeld te zijn. Discriminant-analyse met tal van demografische variabelen levert geen betere voorspellende functie dan op grond van toeval verwacht mag worden. De kenmerken van de niet-effectief behandelde patiënt blijken als volgt geschetst te kunnen worden:

- deze komt vaker van bepaalde psychotherapeuten. Hun effectiviteitspercentage varieert van 53% tot 82%;
- de diagnose heeft weinig invloed;
- kort (< 2 maanden) en lang (6-9 maanden) opgenomen patiënten zijn minder effectief dan de overigen (2-6 maanden en langer dan 9 maanden);
- er is nu geen relatie met het geslacht;
- laag opgeleiden zijn effectiever dan hoog opgeleiden;
- het inkomen houdt geen verband met effectiviteit;
- er is een verschil tussen de intakers, via de ene intaker wordt vaker een niet-effectieve patiënt opgenomen dan een andere;
- niet-effectieve patiënten krijgen na ontslag later voortgezette hulp dan effectieve;
- er is geen verschil in responsratio tussen effectieve en niet-effectieve patiënten.

Een dergelijke evaluatie kan een bijdrage leveren aan beleidsmaatregelen om de effectiviteit te verhogen. De niet-effectieve patiënten zijn immers de doelgroep, waarop de verbetering van de behandeling zich dient te richten.

Hoofdstuk 10. Met verschillende maten gemeten.

Gedurende de periode 1980-1983 is met *meerdere instrumenten* getracht het behandelingsresultaat te meten om na te gaan in welk opzicht deze verschillen. Deze instrumenten waren:

- *Doelevaluaties* van de psychotherapeut, patiënt en externe beoordelaar.
- *Globale beoordelingen* van het effect.
- *Psychologische vragenlijsten* (N.P.V., Zung, M.A.S., F.S.S., I.D.B., S.A., I.W. en Assertiviteit).

De doelevaluaties van de externe beoordelaar zullen later geanalyseerd worden en blijven in dit verslag buiten beschouwing.

In eerste instantie is een *inhoudelijke vergelijking* gemaakt tussen patiëntdoelen, therapeutdoelen en de diagnoses

die de therapeut bij opname stelt. De overeenstemming hiertussen is slechts gering. De doelen welke de psychotherapeut opstelde en maandelijks evalueerde, konden niet geanalyseerd worden: van slechts 50% van de patiënten zijn deze verkregen. Sommige therapeuten bleken deze niet te kunnen of willen formuleren. Van de verkregen doelen ontbraken dikwijls meerdere metingen, zodat het verloop niet meer te traceren was. Inhoudelijk bleken de formuleringen van de operationalisaties in meerderheid te subjectief. Een dergelijke evaluatie-procedure is in de praktijk *niet goed bruikbaar*.

De *standaardvragenlijsten* zijn bij opname, ontslag en 6 maanden follow-up ingevuld door resp. 194, 168 en 111 patiënten. Worden de gemiddelde scores tussen opname-ontslag en opname-follow-up getoetst, dan zijn deze hoegenaamd alle significant (t-toets).

Vervolgens zijn de patiënten in groepen verdeeld naargelang de door de psychotherapeut en henzelf opgegeven diagnoses c.q. problemen. Zo ontstaan groepen van patiënten met resp. dwang, angst, depressie, sociale problemen en, alleen bij de diagnose, persoonlijkheidsstoornissen. Er blijkt dat een sterk verband te bestaan tussen de inhoud van de doelen van de patiënt en de inhoud van de vragenlijsten; zo hebben patiënten die een doel met betrekking tot dwang formuleren, ook een hoge dwang score op de I.D.B. bij opname. De relatie tussen de diagnose van de therapeut en deze lijsten is geringer. De verschillen tussen opname en ontslag bij de psychologische vragenlijsten zijn ook groter bij patiënten met doelen die met de inhoud van de desbetreffende lijsten overeenstemmen. Zo verschilt de score op de I.D.B. significant bij patiënten met een doel dat betrekking heeft op een dwangprobleem, terwijl deze score bij andere patiënten niet of nauwelijks significant verschilt.

Worden vervolgens ook deze effectmaten volgens een criterium geörienteerde methode geëvalueerd, dan blijkt dat slechts een minderheid bij ontslag per vragenlijst een verlaging van de score van hoog naar de bovengrens van normaal laat zien. Het hoogste percentage is steeds de groep patiënten die bij opname en ontslag onveranderlijk hoog scoort.

Deze vergelijkingen tussen de verschillende groepen patiënten, gevormd op basis van de door henzelf opgegeven problemen laat zien, dat zelfs een *homogene groep* als "neurotische" patiënten *differentiatie* behoeft. De groep als geheel evaluerend leidt tot een nivellering van het resultaat.

Tenslotte zijn de verschillende *effectmaten met elkaar vergeleken* om na te gaan in hoeverre deze een overeenkomstige verbetering aangeven. Deze zijn dan: de *verschilscores* tussen opname en ontslag van de psychologische vragenlijsten, de *verbeteringsscore* van de door de patiënt genoemde problemen en de *23 factoren* uit hoofdstuk 9. De correlaties tussen verschilscore en verbeteringsscore zijn hoger dan die tussen verschilscore en de globale beoordeling van het effect. Er is geen verband tussen verschilscore en satisfactie. Opsplitsing

in doelgroepen doet ook hier de bijbehorende correlatie stijgen. Zo is bijv. de correlatie tussen de I.D.B.-verschil score en de verbetering volgens die doelen van de patiënt hoger bij patiënten met een dwangprobleem dan deze correlatie berekend voor de totale groep patiënten.

Geconcludeerd wordt dat de doelen die de patiënt formuleert, goed bruikbare, valide en specifieke effectmetingen zijn, die hoog correleren met andere effectmetingen, en de resultaten doen differentiëren. Het feit dat metingen betreffende satisfactie en beoordelingen van procedures rond het ontslag geen verband vertonen met effectmetingen, maakt het noodzakelijk met *verschillende maten te meten*.

Hoofdstuk 11. Programma-evaluatie geëvalueerd.

Aan het slot van een dergelijk *langlopend evaluatieproject* dat een aantal welomschreven doelen trachtte te bereiken, dienen deze eveneens geëvalueerd te worden.

Het eerste doel: *het ontwerpen van een permanent evaluatiesysteem*, is gerealiseerd, een dergelijk systeem heeft vele jaren in een instelling gefunctioneerd, en er is ervaring opgedaan hoe dit kan worden verbeterd.

Een *vergelijkend onderzoek tussen verschillende meetinstrumenten* levert voldoende gegevens over de concurrente validiteit van met name de zgn. Target-complaints-methode (door patiënt geformuleerde doelen). De hoge betrouwbaarheid maakt inperking van de verzamelde informatie mogelijk.

Dit systeem was bedoeld om *feedback te leveren aan de instelling*. Deze feedback is op diverse niveau's: therapeut, team en directie periodiek verstrekt. Echter, deze heeft *niet* geleid tot een (merkbare) verandering in het beleid. De resultaten bleven geleidelijk dalen. De feedback leidde tot *weerstand*.

Terugkijkend op dit aspect van het evaluatieproject, in het licht van het gegeven dat veel programma-evaluatoren melden dat hun *resultaten niet worden benut*, wordt in hoofdstuk 11 het functioneren van feedback in *organisatieontwikkelingstermen* gezien. Vermoed wordt, dat er in feite geen werkelijke noodzaak tot evaluatie is binnen de G.G.Z.. Het *gebrek aan behoefte* aan of *noodzaak tot feedback* kan worden toegeschreven aan het feit, dat *non-profit organisaties* nauwelijks en niet direct afhankelijk zijn van productiecijfers anders dan aantallen patiënten. Het budget wordt *niet* bepaald door de *kwaliteit van de behandeling*. Tevens bekleden geestelijke gezondheidszorginstellingen een *monopoliepositie*: hun klanten hebben weinig mogelijkheden om naar de concurrent te gaan. Deze *extrinsieke factoren* bewerkstelligen geen veranderingsattitude, zoals dit in *profit-organisaties* wel het geval is.

Indien er ook geen *intrinsieke motivatie* is - op management of medewerkers niveau - om het beleid te veranderen omwille van het belang van de patiënt, zal feedback geen

functie kunnen vervullen; de evaluator zal slechts kunnen constateren dat zijn gegevens niet worden gebruikt. Dit kan bij hem de neiging oproepen om de druk op de organisatie te verhogen, hetgeen tot *weerstandreacties* als *negeren*, *ontkennen of wegverklaren* zal leiden. Uiteindelijk zal de bringer van deze onrustveroorzakende tijding een Socratesiaans lot beschoren kunnen zijn.

Naast dit algemeen geldende probleem van feedback in een non-profit organisatie, zijn aan dit project ook enkele specifieke ervaringsfeiten te ontlenuen, zoals de vermenging van onderzoeks- en programma-evaluatiedoelen en de positie van de evaluator (detachering).

Pro saldo wordt geconstateerd dat de coöperatie van de psychotherapeuten en patiënten over het algemeen goed tot uitstekend is geweest en dat het hoofddoel van dit project is gerealiseerd. Op de afdeling Psychiatrie van het St. Radboud-ziekenhuis worden de ervaringen in extenso benut in het project *Programma-Evaluatie Psychiatrie (P.E.Psy)*. Hierin zal de programma-evaluatie zich uitstrekken tot het curatieve functioneren van polikliniek, dagkliniek en kliniek. P.E.Psy vormt zo een platform van registratie- en evaluatiegegevens, waarop specifieke onderzoeksprojecten van de medewerkers kunnen worden gebaseerd. Zo zal programma-evaluatie de spreekwoordelijke *kloof tussen onderzoek en praktijk kunnen dichten*, maar nu werkend vanuit de praktijkoever.

Met dit resultaat bereikt dit project meer dan zijn beoogde doelstelling. De hier beschreven ervaringen kunnen ook voor andere instellingen in de (G.)G.Z. van nut zijn. *Programma-evaluatie* is een eigentijds instrument waarmee de primaire taak van de eerste inspecteurs van het staatstoezicht op krankzinnigen weer kan worden uitgevoerd: *omzien naar het weldoen aan de (opgenomen) psychiatrische patienten, en vaststellen in hoeverre er wordt welgedaan*. Hiermee is het welzijn van deze patiënten gediend .

Summary

Do well and don't look back" has become an outdated saying in the field of mental health care. Increasingly, the needs for both retrenchment and accountability have given rise to the extrinsic motive of *"looking back"* in order to demonstrate just *"how well"* one has been doing. This tendency - clearly evident in the U.S.A. since the mid-sixties - has led to a growing interest in a new or perhaps modernized form of evaluation research: *"Program Evaluation"*.

By the late seventies this direction of research has come into vogue in the Netherlands under the name of *"Programma Evaluatie"*. This form of applied social research is defined as that field *whose aim is to establish what value a program has by comparing the results obtained with the goals set in advance*. This dissertation gives an account of a program-evaluation project which has been carried out during a period of more than eight years in a clinic for behavior therapy. The description of this project and its results and conclusions is preceded by a detailed theoretical study.

Part A: Program evaluation in theory.

Chapter 1. "Well-doing" in retrospect.

Governmental surveillance as program evaluation.

Program evaluation appears to be a product of the growing need for *accountability* regarding the manner in which government funds are expended. This direction has also been fostered by a promise, not fully justified, of computerized data processing and storage. Chapter 1 examines if and how past Dutch governments have retrospectively evaluated the functioning of community mental health services. This *historical survey* goes back to the founding of the Netherlands as a state.

Under the reign of King William I a beginning was made for improving the care of the insane. These were often indiscriminately locked away in old homes for the lepers along with the homeless, the unemployed and the like. Their housing and care were pitiful. In the period 1800-1841 the situation was dealt with primarily via successive inventories and visits by the Minister of Internal Affairs and Mr. Feith, head of the department concerned. In the late 1830's concrete preparations for legislation were made. These had resulted, in 1841, in the first so-called *Insanity Act* which attempted to lay down rules for the care and treatment of the insane - an Act which may be viewed as having been a blueprint for a goal to be pursued. What the Act has prescribed was, however, far removed from actual practice. In order to check out in how far reality corresponded with the statutory regulations, *the first two inspectors of the state surveillance department*

charged with overseeing the lot of the insane were appointed in 1842: Prof. Dr. J.L.C. Schröder van der Kolk and Mr. C.J. Feith, the inspectorate's head. The first tour of these inspectors - who may be regarded as the first two *program evaluators* in the Netherlands - revealed that only three out of the 28 institutions had met the requirements of the law. They reported as much to the minister in the hope and expectation that he would take remedial measures. However, the country's financial situation did not permit sweeping reforms.

In the years to follow, the inspectors continued to carry out their inspections (the original notes have been preserved) and to make annual reports. These evaluations and feedback have gradually had their effects: the (simple gauge of) *rate of improvement* shows gradual improvement and the number of institutions meeting the requirements of the law rises. Further, the number of custodial institutions (for the so-called untreatable, chronic patients) has declined and the last of which - the "Cellenbroederhuis" in Nijmegen - was shut down in 1886.

One other phenomenon is worth noting, namely, the exponential increase in the number of patients committed: from 1000 in 1845 to almost 5000 in 1885. Naturally this growth was not without consequences for the evaluative function of the inspectors because their number did not increase proportionately. This has led to a shift from *patient-focused* inspection to the more *distant, institution-level*, surveillance. Use was made of the *ratio of recovery* (number of discharged patients/number of patients admitted) as evaluative standard of treatment effectiveness and comparisons across institutions were also made. In 1884 the Act was revised; some of its objectives had become outdated or needed adaptation. Inspection tours continued and the annual reports developed into voluminous tomes. The number of patients increased steadily as well - from the 5000 in 1886 to nearly 26,000 in 1936. However, from 1900 on, the rate of recovery records a decline. From 1909 it was not longer reported. By the modification of the registration in 1930 the recording of the degree of improvement at discharge disappeared - thereby eliminating an evaluative standard at a time when it seemed to be needed most. Such a phenomenon, too, is a common aspect of program evaluation; the security guard departs when someone tries to break in. Finally, in 1936, the stream of annual reports comes to a total halt.

It is not until 1975 that another report (covering the period of 1969 to 1974) was published. By making use of the registration system of the "*Patiënten Intramurale Geestelijke Gezondheidszorg*" (P.I.G.G.), a system established in 1967, the practice of storing national data is reintroduced. In that year, the number of psychiatric patients in public mental hospitals appeared to be almost the same as it was in 1936, namely, 26,000; and in the years to follow it decreased further, reaching 21,000 in 1983. The nature of the reports

is distinctly different from that of the reports of the past century. Apart from concerning itself with clinical activities, the Inspectorate took to overseeing also all newly established residential, semi-institutional and out-patients' facilities for mental health care. State supervision was directed more at the "*input*" and "*effort*" of resources, personal and material, than at the *effect of treatment*. P.I.G.G.-figures, among others, establish that the number of residential patients in public mental hospitals is increasing; and in 1983 there were callings for restoring the use of *custodial institutions* (albeit in a refurbished style) for such patients. In other respects, too, history seems to be repeating itself when once again cases of lengthy separation (isolation) come to attention.

Various lessons relating to program evaluation are to be learned from the history of the State Inspectorate for the Insane:

- Testable (treatment-/result-) *goals* must be formulated.
- These must be *operationalized*.
- There must be safeguards to insure that evaluative information will be used as *feedback for improving results*.
- All participants - government, the evaluated and the evaluating agency - must *cooperate in the interest of improving the evaluation process* itself.

The establishment of a (national) program-evaluation system will give a body to the notion of '*looking back*' on '*well-doing*' which will also undoubtedly promote the welfare of the party most vulnerable - the patients.

Chapter 2. Is evaluation a form of research?

Using De Groot's (1961) definition of research and its manifestations as a point of departure, chapter 2 presents an answer to the (academic) question as to whether program evaluation can be regarded as a form of social science research. De Groot considers the "*evaluation problem*" including the evaluation of the effects of treatment as being the basic model which underlies hypothesis testing. Not without misgivings is the assumption in what follows being made that this refers to the experimental testing of hypotheses.

Hofstee (1980) - who defines research along the lines of the so-called *wager-model* (verifying who is *right*) - also attaches considerable significance to the evaluation of effects. He warns, however, that applied research is often used for "*public relations*" purposes and that, as a research investment, it is socially irrelevant because the party being investigated would not wish to admit being in the wrong.

The difference between *research* and *evaluation* lies primarily in the nature of the information gained and the purpose of the scientific activity. Research is essentially *nomothetic*, being concerned basically with acquiring *knowledge* and

uncovering lawful patterns. Evaluation, on the other hand, is basically *idiographic* in nature and is meant to demonstrate the *capabilities*. This vision implies a distinction involves a shift from the primary to the secondary target (increasing scientific knowledge versus improving treatment outcome). From the perspective of methodology, this has the consequences of *reversing the dependent and independent variable*; and, further, it affects both the testing procedure (normative or absolute) and the interpretation of the results (statistically significant or clinically relevant).

Treatment evaluation may manifest itself in many forms: *summative or formative and process or outcome evaluation*. Program evaluation is a form of treatment evaluation wherein primarily outcome or effect information is being gathered and the results are used for formative purposes particularly on an institutional level. It may be concluded that the discussion among social scientists as to whether program evaluation is a form of scientific research is irrelevant and pointless.

Chapter 3. Program evaluation: what's that?

Program evaluation may be defined as follows: *determining the value of a program by comparing its effects with its (previously defined) aims*. A program is any activity (generally sponsored from public funds) aimed at improving the social, economic or psychological welfare of the target populations. The *goal of program evaluation is to provide feedback* for improving the functioning of the program.

Program evaluation may be directed at any of the various aspects of a program: *effort* (budget, personnel, target group, etc.); *performance* (effects, satisfaction, etc.); *adequacy* (relationship between results and needs); *efficiency* (relationship between effort and achievement); and, *process* (why these results have been achieved). This last aspect, it may be remarked, is usually considered to fall beyond the domain of program evaluation.

Evaluators can or rather should *restrict themselves to giving feedback*, describing the status quo in relation to the objectives. As to the question of "*why?*", the evaluators will at best have indirect and too little information for answering it (adequately). In view of the differing concerns of mental health institutions on the one hand and government/funding party on the other, a distinction is made between *internal* program evaluation - carried out by the institution itself - and *external* program evaluation - carried out by (and for) an external organ. One other classification is used, and that is for the level at which program evaluation is focused: patient, program and institution or group of institutions (regional or national). With respect to program evaluation procedure, the nature (internal/external) or level (patient/national) are of little consequence. The procedure of program evaluation is a *cyclical process*:

- The goals of the program as well as the concrete or objective needs for information must be specified for each policy level in a *policy plan*.
- Next comes the articulation of an *evaluation plan* in which this need for information is operationally defined by means of measuring instruments and clear procedures such as who is going to supply what information about whom.
- *Implementation* of the plan takes place in stages preferably in accordance with the so-called *onion-model*, wherein one proceeds stepwise from the outermost region inward. Evaluation of intake/hospitalization takes place first and this may be followed by internal, interim evaluations possibly involving the gathering of treatment data.
- The final stage of the first cycle pertains to *evaluation and feedback*. The data which have been collected are analyzed, their relations to the targets examined and finally reported to the parties concerned. This stage leaves open the possibility for modifying the policy plan and starting the evaluation cycle anew.

This feedback will not always be received favorably. Program evaluators complain that the data they supply are insufficiently utilized for they provoke *resistance reactions* of which three are common: *explaining away, denying, and bombarding the evaluator with a barrage of questions*. Program evaluation is a mirror through which an institution looks at itself; the management has the task of addressing the possible imperfections which the feedback brings to light.

Chapter 4. Measuring the effects of psychotherapy.

One of the major problems of carrying out psychotherapy research is the difficulty of *operationalization*. Through the years a wide range of instruments has been developed and utilized in (experimental) comparative research designed for measuring the effects of the various therapies. These instruments vary from projective techniques to questionnaires, from behavioral observations to global ratings of outcome and can be appraised on such dimensions as reactivity and specificity in addition to the usual psychometric parameters. With the recommendations of several extensive meta-evaluative studies in mind, three methods of evaluation are described in detail: *target evaluations, psychological questionnaires and subjective effect/satisfaction judgment*.

Of the individualized target evaluation methods, two procedures are widely used: "*Target Complaints*" and "*Goal Attainment Scaling*" (G.A.S.). The first of these two is especially favored by researchers; its administration is economical and its use scarcely requires professional preparation. Use of the G.A.S. is time-consuming and demands constant attention, aid and guidance; consequently, it is not too popular among researchers and its simplified version, the *Individual Target Evaluation (Measure)* seems to offer greater potential.

The instruments used most frequently in evaluation are the *psychological questionnaires*. Close inspection indicates that the questionnaires most often used in the Netherlands scarcely meet accepted standards of reliability and validity-qualities which, along with the ease of use, were the main reasons for having been chosen in the first place. Such questionnaires do, however, offer the advantage of yielding comparable information across studies.

Subjective judgements of therapeutic results and the degree of satisfaction with treatment have probably been used most. By means of ad hoc instruments patients are asked to *judge the results obtained* and to indicate the extent of their *satisfaction* with what has been offered. Since a large variety of these so-called "home-made" instruments is in use, information of comparative nature is very difficult to attain. An instrument tested in numerous studies and refined through various analyses is the *Client Satisfaction Questionnaire (C.S.Q.)*. This questionnaire, which has been translated also into Dutch, seems to be quite suited for evaluation when used in conjunction with other data such as treatment results.

Preferably, the evaluation of psychotherapeutic results should be carried out by means of a "*multitrait-multimethod*" approach. Inasmuch as psychotherapy can have multiple and different effects, it is desirable to obtain an assessment of both *direct therapeutic results* (target measures) and patients' subjective evaluation of *satisfaction*. In the long run such an approach has the merit of yielding comparative data about differing therapeutic methods - thereby contributing to the identification of (causal) connections between the therapeutic goals and the means used for their attainment.

Part B. Program evaluation in practice.

Chapter 5. The site of the research, the Overwaal Clinic.

A program evaluation project has been carried out in a psychotherapeutic clinic (30 beds) for a period spanning more than eight years. Chapter 5 describes the features of this site of the study, the Overwaal Clinic.

- The treatment strategy is primarily *behavior therapy*.
- Treatment is highly *individualized*. In cooperation with a sociotherapist, a psychotherapist implements an individually tailored plan of treatment. Group therapies such as those involving discussions, body movement and creative activities are also used.
- Treatment is *intensive* both in terms of staff participation and time. There is a psychotherapist and sociotherapist (B-level nurse) for every 4-5 patients and the assistance of movement- en creativity-therapists is also employed when needed.

- Treatment is *short-term*, meeting as much as possible the so-called A1-standard (i.e., less than six months).
- Treatment is geared to *neurotic adult patients* who cannot (or can no longer) be treated as out-patients.
- In principle this intensive clinical intervention is succeeded (again) by treatment on an *out-patient basis*; one of the goals of the Clinic is to make it possible for patients to put out-patient services to fruitful use.

Chapter 6. The Overwaal Program-Evaluation Project.

Upon its request, a clinical/research psychologist was detached in 1976 to the Overwaal Clinic by the Department of Psychiatry of the Catholic University of Nijmegen. The purpose of this appointment was the setting up of an *evaluative system* which would provide *internal feedback* on the results of therapy. The Overwaal Program-Evaluation Project, carried out by this psychologist, may be divided into four chronological phases:

- A *retrospective follow-up study* of the period 1974-1976. This allowed the psychologist to gain experience with the discharge and follow-up questionnaires to be used.
- In the period of 1976-1980 a *permanent evaluation study* was undertaken in which, at the time of both hospitalization and discharge and at three follow-up periods, patients and therapists supplied information about numerous aspects of the course and results of therapy.
- From 1980 to 1983 a *comparative investigation* of various measuring instruments was conducted: *target evaluations* by the patient, the therapist and an external judge; *standard psychological questionnaires*; and *global judgments of effect and satisfaction*. (Financial support for this phase of the study was provided by the *Prevention Fund* for additional personnel).
- In 1983 the range of data to be collected was reduced. As a result, only the patients' target evaluations, a few psychological questionnaires and the global effect/satisfaction judgements were continued.

At the end of 1984 the appointment of the psychologist detached to the Clinic was terminated, thereby bringing the Project to a partial halt. In 1985 the project was allowed to continue administratively and a year later it was taken over by a psychotherapist at Overwaal. The psychotherapist curtailed the amount of data to be gathered and altered the format of its collection, a step which has put a final end to the Overwaal Program-Evaluation Project. This research report covers the data processed between 1974 and mid-1985.

Chapter 7. The quantitative data on record.

The course of an evaluation project proceeds from *data recording to evaluation*. The setting up of an information system is essential for allowing one to get a good description of the more objective patient data. Such data, pertaining to more than 800 Overwaal patients, are presented in chapter 7. The data are expressly grouped into periods of two or three years in order to facilitate the detection of possible trends.

Looking at the entire group of patients, the following descriptive picture emerges:

- Approximately 80 patients are hospitalized each year. Their *average stay* is 140 days with a standard deviation of 84 days. Treatment of about 20% of the patients is terminated within two months from admission; most of these may be construed as "drop-outs". (This percentage has increased over the years). About 25% of the patients were hospitalized in excess of the AI-standard of six months.

- With regard to *demographic variables*, 60% of the Overwaal patients are females and 40% are males. This ratio has varied from 50-50 to 67-33 over the years. The average age is well into the thirties (SD = 9); more than 60% are married/cohabitants; about 55% fall on the lower range of the socioeconomic scale. The latest educational attainment of 50% is primary school, lower vocational and mid-level secondary school. Prior to hospitalization, the main source of income was a government disability insurance (37%), employment (25%), nearly 25% had no personal income of their own (housewife/husband).

- *The nature and duration of the problems* are determined on the basis of both diagnostic intake which precedes hospitalization and patients' own statements as to when they have become conscious of their problems and how long they have been receiving treatment to alleviate them. About 60% of the patients are diagnosed as *neurotic*, obsessional-compulsions and phobias being the most common manifestations. The durations of the complaints exceeds three years and these have often led to long periods of dysfunctioning. More than half of the patients have received treatment spanning a period of over three years and over 25% have been treated for six or more years. The main sources of referral are the family physician or the "Regional Institute for Mental Health" (RIAGG). Inspection of pre-hospitalization care reveals that 40% of the patients have been previously hospitalized and that 90% have been treated before on an out-patient basis. Treatment was initiated on the average 42 months prior to admission to Overwaal; however, 20 of these months have gone without treatment - implying a strong history of treatment interruptions.

The data presented thus far are useful, first, for checking the correspondence between the *actual population* and the *target group* and, second, for exploring the relationships

between these variables and effect measures - an issue to be taken up in chapter 9.

Chapter 8. Evaluation of the results.

The results of three consecutive evaluation periods are described separately; the data could not judiciously be pooled because of the alterations made in the measuring instruments used.

In the *retrospective follow-up investigation (1974-1976)* information was obtained from 125 patients (a response rate of 70%). Forty percent of the respondents indicated that they considered their problems as having completely/almost completely disappeared; 45% as having changed little; and, 15% reported no improvement. Although these results are not impressive, patients did, however, report a relatively high degree of satisfaction with hospitalization and treatment. There was a significant difference in both effect and satisfaction between patients who had been treated by the permanent staff on the one hand and those treated by (supervized) resident psychiatrists. Due to artifacts in questions, heterogeneity of the patients and the small size of the group these data were not interpreted further.

On the basis of the experiences which the investigator has gained using this questionnaire (of his own construction) a *permanent evaluation*, beginning in the autumn of 1976, was carried out. The rate of response to the questionnaires, which were either handed out or mailed, ranged from high to very high (71-100%). This percentage had dropped as the follow-up duration grew longer.

The most important evaluative data are:

- *Decreasing effectiveness of treatment.* In 1977, 31% of the patients (N= 262) reported upon discharge that their problems were a bit improved/not improved/gotten worse. In 1980 this percentage has increased to 53. (It may be pointed out that responses to other relevant items of the questionnaire reveal the same pattern).

- Psychotherapists, too, have indicated a *decline* - albeit lower - in the treatment *results*.

- *Satisfaction* with treatment is *high*.

- At the *first follow-up* there appeared to be a *sharp drop* in the patients' evaluation of their condition; while at discharge 25% said that "it is not going so well" or "it is bad", 40% said so at the first follow-up. In the course of the period succeeding discharge the rate fell to 31%. In the follow-up the decline in results is also evident.

- *Staff evaluations* of patients' status quo after discharge disclosed a similar pattern; in 1977, 63% characterized hospitalization as "meaningful" but in 1980 it was only 41%.

Summing up, the data reveal a consistent trend of decline in the reported effectiveness of treatment. This information,

being as it is a product of a descriptive evaluation project can fulfill a feedback function - the goal of an evaluation project of this genre.

In 1980 the questionnaires have been drastically revised though the basic design remained intact. A noteworthy extension pertains to the patient's individualized target evaluation (Target-complaints-method). The response rate remained unaltered or attained a higher level (72-100%) than was the case in the period which preceded.

The most important results of this evaluation period are:

- *Treatment results* (as measured by various questions) at discharge again showed a downward trend. An exception to this occurred in 1984 where the deviation was in the positive direction. All in all, about 40% of the evaluations proved to be positive, about 38% neutral and the remaining 30% negative (N = 320). Differences were evident in every evaluation measure used.

- On the whole patients' *satisfaction* was high and the evaluations of psycho/sociotherapists were also positive.

- Therapists' evaluations were more *favorable* than those of patients. The prognosis - as to how the condition of the patient would fare as time goes by after discharge - was less favorable than the initial evaluation, and patients' condition did indeed decline.

- In the *follow-up period* a phenomenon similar to that observed for the years 1976-1980 was also noted: evaluation of the status quo declined strongly during the first follow-up. In the course of time, however, a rise was discerned but without ever reaching the level of evaluation at discharge. In this respect, too, 1984 showed a more favorable picture than might be expected in light of the tendency since 1981.

- The evaluations made by *therapists after discharge* correspond with those of the patients except for being generally more positive.

A comparison of the responses to questions pertaining to the two periods of evaluation, 1976-1980 and 1981-1985, showed that the slight changes in the formulation of the answer categories may have led to big shifts. The newer questionnaires offered a better gradation of response possibilities than the earlier versions.

Despite these differences, the results over the total period of more than eight years may cautiously be summarized as follows:

- An increasing percentage of patients indicate minor therapy results; the rise from approximately 20% in 1977 to 39% in 1985 suggests that the problem associated with hospitalization may have been insufficiently improved, if at all, or that deterioration may have set in. While in 1977, 19% reported that their situation was "bad" or "not very good" the figure grew to 31% in 1985.

- Follow-up results project a similar picture. The first follow-up showed the least favorable results: in 1977, 49% maintained that things were "not going well" or that they were "bad"; in 1983, this percentage rose to 52% and in 1984 it declined again to 28%.

- Satisfaction with treatment was constantly high. About 70-90% of patients reported being (very) satisfied with the various aspects of treatment including the therapists.

- For 80% of the patients arrangement for after-care was made three weeks prior to discharge. (Therapists after discharge spoke of a corresponding figure of 60%). After-care arrangements made too late create a hiatus in the curative process. Only in the case of roughly 30% of the patients was the transition direct between treatment received at the Clinic and after-care.

- The follow-up data reveal that the problems had apparently grown worse immediately after discharge. However, a gradual improvement was noted for the ensuing period as revealed in the 6-12 months' follow-up.

- One year after discharge, 30% of the patients were able to function without further help; the remaining 70% were in need of help or liked to receive it.

- Due to want of explicitly defined goals, the evaluations of the Overwaal program is *more descriptive than is optimal*. Nonetheless this chapter has outlined the range of information which such a project can provide.

Chapter 9. Analysis of the results.

The immensely rich data of the Overwaal Program-Evaluation Project (approximately 700 variables for each of the circa 800 patients) entices further analysis. This analysis, presented in chapter 9, is mainly focused on the situation since 1980. *Reliability* was examined first yielding high indices for both internal consistency and inter-rate agreement.

Data based on the discharge and follow-up questionnaires were narrowed down to 23 factors by means of *factor analysis*. The loading of the questions pertaining to the factors were consistently high (.50-.90) with an explained variance of circa 60% for each questionnaire. The 23 factors have been used as dependent variables in order to examine their relationships to diverse independent variables. *Multivariate analysis of variance* has been applied for testing the existence of differences between psychotherapists, diagnosis at hospitalization and year of discharge. Of these main effects, the variables of the psychotherapist in particular showed a significant difference in the therapy effectiveness variable and in psychotherapist's evaluation of the patient at discharge and follow-up. Of all diagnostic groups only the one diagnosed as "personality disorder" showed a significant difference on the effectiveness variable in the therapist's discharge questionnaire. Contamination cannot be ruled out, however, because the evaluation of effectiveness by the

patients themselves who were so diagnosed failed to reveal a corresponding difference.

Numerous demographic patient variables have been investigated by means of *univariate analysis of variance* to check if they were related to differences in effect. To some extent, this appeared to be the case with the variable of gender (larger effect among males than females) and arrangement/start of after-care; such help starting too late after discharge is related to a smaller amount of effect. There appears to be scarcely a correlation between the duration of the complaints on the one hand and help intervention prior to hospitalization on the other.

Finally, a *criterion-oriented analysis* was carried out on the therapy results. Patients were dichotomized into effectively treated/ineffectively treated. Therapy was found to have been effective in the case of 66% of the entire group. Discriminant analysis with numerous demographic variables yielded no better predictive function than chance. The characteristics of the "not effectively treated" patients may be outlined as follows:

- Frequently these have been treated by specific, identifiable therapists whose effectiveness rate, nonetheless, ranged from 53 to 82%.
- Diagnosis appears to be of little consequence.
- Patients not effectively treated were hospitalized for a short period (less than 2 months) or for a period of 6-9 months; those hospitalized for 2-6 months or longer than 9 months fared better.
- Gender was irrelevant.
- Patients with hardly any schooling were treated more effectively than those with better education.
- Income was unrelated to effectiveness.
- There was an intaker effect; some particular intakers have frequently handled more of the "not effectively treated" cases than other intake colleagues.
- After-care initiated too long after discharge was more common among cases treated ineffectively.
- The response rate was unrelated to treatment effectiveness.

Evaluations of the sort just outlined have the potential of contributing to policy measures aimed at improving effectiveness. Patients ineffectively treated do form a particular challenge and are a target group on whom effort for improving the rate of treatment outcome should be focused.

Chapter 10. The measuring instruments considered.

During the period of 1980-1983 numerous instruments for measuring therapeutic results were used and their utility compared. These instruments were:

- *Target evaluation* by the psychotherapist, patient and by external judge.
- *Global judgements of effect*.

- *Standard psychological questionnaires* (N.P.V., Zung, M.A.S., F.S.S., I.D.B., S.A., I.W. and Assertiveness).

Analysis of the target evaluations by external judge will be analyzed at a later date and is thus beyond the scope of this report.

The first step in this undertaking involved a *comparison of the content* of the targets formulated by both patient and therapist and of the diagnosis at admission. Agreement among these sources was quite limited. The targets which the therapist had initially set and then evaluated monthly could not be analyzed since data for only 50% of the patients were available. Further, it appeared that some therapists could not, or did not want to, formulate the targets. Often, too, no assessments were made of the targets attained and, consequently, their course could not be traced. Finally, it appeared that most of the operationalizations were cast in terms too subjective to permit utilization. All in all, such an evaluation procedure is of *little practical use*.

At hospitalization, discharge and the 6-months' follow-up the *standard questionnaires* had been filled in by respectively 194, 168 and 111 patients. The difference in the average scores between hospitalization-discharge and hospitalization-follow-up were tested (t-test) and nearly all were found to be statistically *significant*.

Patients were then classified into groups on the basis of the presenting complaints and diagnosis. This yielded groups with compulsions, anxiety, depression and social problems and, on grounds of merely diagnosis, personality disorders syndrome. There appeared to be a stronger relationship between the contents of the patients' targets and the contents of the questionnaires (patients who formulated a target concerning compulsion, for example, also had a high "I.D.B." compulsion score at admission) than between these questionnaires and the therapists' diagnosis. The difference between psychological questionnaires filled in at admission and discharge were also greater for those patients whose targets corresponded with the contents of the respective lists. For example, the "I.D.B." score differed significantly among patients with a target regarding a compulsion problem while this score hardly differed among other patients.

When these effect measures were evaluated along the lines of a criterion-oriented method it appeared that only a minority per questionnaire showed, at discharge a decrease in score from high to the upper limits of normal. The highest percentage was invariably among the group of patients who consistently had a high score at admission and discharge.

These comparisons (of patients grouped on the basis of their own statements of the problem) showed that even a relatively *homogeneous group* such as "neurotic" patients requires *differentiation*. Evaluating the group as a whole leads to a leveling of result.

Finally, a comparison was made of the various effect measures in order to find out the extent to which these reflect a corresponding improvement. These consist of *difference in score* on the psychological questionnaires filled out at admission and discharge, the *improvement score* on the problems specified by the patients and the *23 factors* discussed in chapter 9. The correlations between difference- and improvement-score were higher than those between difference-score and the judgment of effect. There was no relationship between difference-score and satisfaction. Here, too, splitting up the groups improved the size of correlation. For example, the correlation between the I.D.B.-difference score and the effect was higher among patients with a compulsion problem than the correlation computed for the total group of patients.

The conclusion is warranted that the targets which patients formulate do have utility, are valid and are bona fide effect measures which correlate highly with other effect measures and bring about refined differentiations in results. The fact that measures having to do with satisfaction and evaluation of procedures at discharge showed no relationship to effect measures, dictates *the need for the use of differing assessment measures*.

Chapter 11. Program-evaluation evaluated.

At the end of such a lengthy evaluation project which aimed at attaining a few carefully described goals these goals, too, should be evaluated.

The first goal of *designing a permanent evaluation system has been realized*. Such a system has functioned for many years in an institution and experience has been gained in how it can be improved.

A comparative study of different measuring or assessment instruments has yielded adequate data on the concurrent validity of particularly the so-called target-complaints-method (targets being formulated by the patients). Its high level of reliability makes possible the curtailment of information gathered.

This system was designed *to give feedback to the institution*. This feedback was given periodically at different levels: therapist, team and management. It did not, however, result in a (noticeable) modification of policy. The results kept declining, *the feedback was ignored and resistance had set in*, possibly generating one of the motives which led to the termination of the project.

Mulling over this aspect of the evaluation project and in the light of the fact that many program evaluators complain that *their results are not optimally utilized*, a sketch of how feedback functions in *organizational development terms* was presented in chapter 11. One suspects that there was no real or genuine need for evaluation beyond the purpose of

public relations. The fact that the objectives of the institution were not (or could not be) made explicit in addition to the manner in which feedback was dealt with, are signs of this. *The lack of desire or need for feedback* can be ascribed to the fact that non-profit organizations are scarcely and not directly dependent for sustenance upon data beyond patient numbers. *The budget is not determined by the quality of the treatment given.* Furthermore, mental health agencies are in a *monopoly position*: their clients are nearly captive with hardly the possibility of knocking on the doors of competitors. These *extrinsic factors* do not generate the sort of attitude change one encounters in profit-making organizations.

And so long as the *intrinsic motive*, on the level of management and staff, for policy change in the interest of patients is absent, feedback cannot realize any purpose and evaluators are assured that their data are not being put to use. Evaluators who press the organization are bound to engender *reactions of resistance such as ignoring, denying and explaining away* and their disturbing tidings will eventually hasten to evaluators the *fate of Socrates*.

Apart from highlighting the general feedback problem in a non-profit agency experiences in this project brought to the fore other specific, problematic issues such as the mixing of research and program evaluations targets and the position of the (detached) evaluator.

All clouds have a silver lining. One such lining is the cooperation of the patients and therapists which in general ranged from good to excellent. And there was also the realization of the main objective of the project. The Department of Psychiatry of the St. Radboud Hospital has planned to use the findings quite optimally in its own program evaluation project (*Programma-Evaluatie Psychiatrie, P.E.Psy*). In this project program evaluation will encompass the curative functioning of the out-patient clinic, day hospital and clinic. The P.E.Psy constitutes a platform for recording and evaluating data upon which specific research projects of staff members can be based. Program evaluation, as a field of endeavor, may finally succeed in *bridging the gap between research and practice* but building now from the practice-side.

With such a development the project has achieved more than was intended. The experiences described here can also be of use to other (mental) health care agencies. *Program evaluation* is a contemporary instrument through which the primary task of the first inspectors of the State Inspectorate for the Insane can be carried out again: *"Looking back" at the "well-doing" for (hospitalized) psychiatric patients, and establishing just how well this "well-doing" is in reality being done.* So doing will promote the welfare of these patients.

Bibliografie

- Aarsen, R. (1984). Met het oog op kwaliteit: een pleidooi voor interne evaluatie. In: Schippers, e.a. *Programma-evaluatie in de geestelijke gezondheidszorg*. Swets en Zeitlinger, Lisse.
- Anderson. (1978). *The profession and practice of program evaluation*. Sage Publications, London.
- Arrindell, W.A., Ettema, M. (1978). De symptom checklist (S.C.L.-90): dimensionele structuur, betrouwbaarheid en validiteit van een klachtenlijst in een normale populatie. *Heymans Bulletin*, Psych. Inst. R.U.G. (HB-78-768-EX).
- Arrindell, W.A. (1980). Dimensional structure and psychopathology correlates of the Fear Survey Schedule (F.S.S. III) in a phobic population: a factorial definition of agoraphobia. *Behavior research and therapy*, 18, pag. 229-242.
- Arrindell, W.A., Ettema, M. (1981). Dimensionele structuur, betrouwbaarheid en validiteit van de Nederlandse bewerking van de symptom checklist (S.C.L.-90); gegevens gebaseerd op een fobische en een normale populatie. *Nederlands tijdschrift voor de psychologie*, 36, pag. 77-108.
- Arrindell, W.A. & Ende, J. van der (1986). Further evidence for cross-sample invariance of phobic factors: psychiatric inpatient ratings on the fear survey schedule III. *Beh. res. ther.* 24,3 pag. 289-297.
- Attkisson, C. (e.a.) (1974). A working model for mental health program evaluation. *American journal of orthopsychiatry*, 44 (5), pag. 741-753.
- Attkisson, C., Hargreaves, W.A., e.d. (ed.) (1978). *Evaluation of human service programs*. Academic press, New York.
- Attkisson, C. en Broskroski (1978). Evaluation and the emerging human service concept. In: Attkisson, C., Hargreaves, W.A., e.d. (ed.) *Evaluation of human service programs*. Academic press, New York.
- Attkisson, C.C., Roberts, R.E., Pascoe, G.C. (1983). The evaluation ranking scale: clarification of methodological and procedural issues. *Evaluation and program planning*, 6 (3,4), pag. 349-358.
- Bandura, A. (1969). *Principles of behavior modification*. Holt, Rinehart & Winston, New York.
- Barlow, D.H., Hayes, S.C., Nelson, R.O. (1984). *The scientist practitioner*. Pergamon Press, New York.
- Battle, C.C. e.a. (1966). Targets complaints as criteria of improvement. *American Journal of psychotherapy*. 20, pag. 184-192.
- Bayer, H.A.P.M. (1984). De betekenis van registratie voor het beleid. *Maandblad geestelijke volksgezondheid*, 3, pag. 267-276.
- Beenackers, A.A.J.M., Donker, M.C.H. (1981). *Evaluatie crisis centrum Utrecht*. Rapport V, eindevaluatie.

- Beenackers, A. (1984a). Programma evaluatie t.b.v. externe beleidsinstanties en uitgevoerd door externe onderzoekers. In: Schippers, e.a. *Programma-evaluatie in de geestelijke gezondheidszorg*. Swets en Zeitlinger, Lisse.
- Beenackers, A. (1984b). Weerstanden en verweer: het spanning-sveld tussen onderzoeker en onderzochten. In: Schippers, e.a. *Programma-evaluatie in de geestelijke gezondheidszorg*. Swets en Zeitlinger, Lisse.
- Bennis, W.G., Benne, K.D. & Chin, R. (1969). *The planning of Change*. Holt, Rinehart & Winston, New York.
- Bergin, A.E., Lambert, M.J. (1978). The evaluation of therapeutic outcomes. In: Garfield, S.L., Bergin, A.E. *Handbook of psychotherapy and behavior change*. John Wiley & Sons, New York.
- Bijbelgenootschap, Het Nederlandse (1974). *Het oude testament*. Amsterdam.
- Binnenveld, H. (1985). *Filantropie, repressie en medische zorg. Geschiedenis van de inrichtingspsychiatrie*. Van Loghum Slaterus. Deventer.
- Blake, R.R. & Mouton, J.S. (1964). *The manegirial grid*. Gulf Publishing Comp. Houston, Texas.
- Blake, R.R. & Mouton, J.S. (1977 vert. M. Zwiers). *Leiderschapspatronen*. Het Spectrum, Utrecht.
- Bloom, B.S., Hastings, J.T., Madaus, G.F. (1971). *Handbook of formative and summative evaluation of student learning*. McGraw-Hill Book Company, New York.
- Blöte, A.W., Curfs, L.M.G., Foran, J.C. (1983). Evaluatie-onderzoek en de persoonlijke factor. *De psycholoog*, 18 (10), pag. 559-567.
- Bonninghausen, L.D.T.M. (z.j.). *F.S.S. Intern rapport nr. 4*, I.C.I.P., Rijksuniversiteit, Utrecht.
- Bonninghausen, L.D.T.M. (z.j.). *Manifest anxiety scale*. Intern rapport nr. 2, I.C.I.P., Rijksuniversiteit, Utrecht.
- Bonstedt, Th. (1973). Concrete goal-setting for patients in a day-hospital. *Evaluation*, 1 (2). Mental Health Center.
- Brey, H. de (1981). *Richtlijnen voor programma-evaluatie van de Evaluation Research Society*. Stichting Overlegorgaan GGZ, Rotterdam.
- Brey, H. de, (1983). A cross-national validation of the client satisfaction questionnaire: the Dutch experience. *Evaluation and program planning*, 6 (3,4), pag. 395-400.
- Brinkman, W. (1977). *Een assertiviteitsschaal* Intern rapport afd. groepspsychologie, psychol. lab., G.U. Amsterdam.
- Brinkman, W. (1978). Het gedragstherapeutisch proces. *Handboek Gedragstherapie* Van Loghum Slaterus, Deventer.
- Brintnall, J., Garwick, G. (1976). *Applications of goal attainment scaling*.
- Broekman, T., Schaap, C. (1984). Programma-evaluatie in twee instellingen voor psychosociale hulpverlening. In: Schippers, G.M. (ed.) *Programma-evaluatie in de geestelijke gezondheidszorg*. Swets en Zeitlinger, Lisse.

- Brook, O.H. (1985). "New long-stay"patiënten in de algemene psychiatrische ziekenhuizen. Probleem van zorg en zorg-probleem van het psychiatrische ziekenhuis. *Het ziekenhuis*. 15, pag. 98-100.
- Broskowski, A. & Driscoll, J. (1978). The organisational context of program evaluation. In: Attkisson, C.C. & Hargreaves, W.A. *Evaluation of human service programs*. Academic Press, New York.
- Bruyn, E.E.J. de (1985). *Psychodiagnostiek*. Lemniscaat, Rotterdam.
- Butters, E. (e.a.) (1977). *The Denver Community Mental Health Questionnaire: replication of the psychometric analysis and addition of the home productivity scale*.
- Campbell D.T. & Fiske, D.W. (1959). Convergent and discriminant validity by the multitrait-multimethod matrix. *Psychol. Bull.* 56, pag. 81-105.
- Campbell D.T. & Stanley, J.C. (1963) Experimental and Quasi-experimental Designs for research on teaching. In: Gage, N.L. (ed.) *Handbook of Research on Teaching*. Rand Mc Nally Company, Chicago.
- Ciarlo, J., Davis, C. (1976). *Second and third quarter 1975 outcomes for adult clients*. Northwest Denver.
- Ciarlo, J., Reihman, J. (1978). *The Denver Community Mental Health Questionnaire: development of a multi-dimensional program evaluation instrument I*. Mental health systems evaluation project of the northwest Denver Mental Health Center and the University of Denver.
- Ciarlo, J.A. (1981). *Utilizing evaluation (concepts and measurements techniques)*. Sage Publications, London.
- Classens, T., Gommans, J. (1983). *Vragenlijst Algemeen Welbevinden. Een screeningsinstrument voor psychiatrische problematiek?* Scriptie examen psychologisch medewerkster (N.I.P.) en stageverslag sociologie van het gezin en sociale welzijnszorg. Katholieke Universiteit, Nijmegen.
- Cook, T.D. & Campbell, D.T. (1979) *Quasi-Experimentation*. Rand Mc Nally College Publishing Comp. Chicago.
- Corveleyn, J. e.a. (1978) Goal Attainment Scaling bij chronische psychotische patiënten in het raam van een medicamenteus onderzoek. In: *Maandblad Geestelijke Volksgezondheid*, 33,4 p. 262-276.
- Coursey, R.D. (ed.) (1977). *Program evaluation for mental health*. Grune & Stratton, New York
- Cronbach, L.J. e.a. (1980). *Toward reform of programm evaluation* Jossey-Bass Publ. San Francisco
- Datta, L., Perloff, R. (1979). *Improving evaluations*. Sage Publications, London.
- Davis, H.R. (1973). Four ways to goal attainment. *Evaluation*, 1, pag. 23-28.
- Dekker, H.M. (1984). Programma-evaluatie en overheidsbeleid ten aanzien van de GGZ. In: Schippers, e.a. *Programma-evaluatie in de geestelijke gezondheidszorg*. Swets en Zeitlinger, Lisse.

- Derogatis, L.R., Lipman, R.S., Covi, L. (1973). The S.C.L.-90: an outpatient psychiatric rating scale. *Psychopharmacological bulletin*, 9, pag. 13-28.
- Derogatis, L.R., Cleary, P.A. (1977a). Confirmation of the dimensional structure of the S.C.L.-90: a study in construct validation. *Journal of clinical psychology*, 33, pag. 981-989.
- Derogatis, L.R. (1977b). *S.C.L.-90: administration, scoring en procedures manual - I for the revised version*. John Hopkins University School of Medicine, clinical psychometrics research unit, Baltimore.
- Descartes, R. (1977, vert. Th. Verbeek). *Over de methode*. Boom, Meppel.
- Deutsch, M., Krauss, R.M. (1965). *Theories in social psychology*. Basic Books Inc., New York.
- Dijkhuis, J.H. (1960) *Het beoordelen in de psychologie*. Proefschrift Universiteit, Utrecht.
- Dijkstra, P. (1974). De zelfbeoordelingsschaal voor depressie van Zung. In: Van Praag, H.M. van & Rooijmans, H.G.M. *Stemming en ontstemming*. Bohn, Amsterdam.
- Donker, M. (1984a). *Het profielpakket: overzicht van de mogelijkheden*. N.c.G.v., Utrecht.
- Donker, M. (1984b). Inleiding op de forumdiscussie. In: Schippers, e.a. *Programma-evaluatie in de geestelijke gezondheidszorg*. Swets en Zeitlinger, Lisse.
- Drenth, P.J.D. (1972). *De psychologische test*. Van Loghum Slaterus, Deventer.
- Ellis, R.H., Wilson, N.C.Z. (1973). Evaluation treatment effectiveness using a goal-oriented automated progress note. *Evaluation*, 1, pag. 6-11.
- Ellmer, R., Olbrisch, M.E. (1983). The contribution of a cultural perspective in understanding and evaluating client satisfaction. *Evaluation and program planning*, 6 (3,4), pag. 275-282.
- Endecott, R.L. e.a. (1976). The global assessment scale: A procedure for measurement overall severity of psychiatric disturbance. *Archives of General Psychiatry* 33, pag. 766-771.
- Esch, P. van der (z.j.) *Geschiedenis van het Staatstoezicht op krankzinnigen. Deel I, II, III*. Staatsuitgeverij. 's-Gravenhage.
- Eysenck, B.G., Eysenck H.J. (1964). An improved short questionnaire for the measurement of extraversion and neuroticism. *Life sciences*, 3, pag. 1103-1109.
- Eysenck, H.J., Eysenck, S.B.G. (1964). *Manual of the Eysenck personality inventory*. Hodder & Stoughton, London.
- Fink, A., Kosecoff, J. (1980). *An evaluation primer and workbook (practical exercises for health professionals)*. Sage Publications, London.
- Fiske, D.W., Hunt, H.F., Luborsky, L. (e.a.) (1970). Planning of research on effectiveness of psychotherapy. *Archives of general psychiatry*, 22, pag. 22-32.

- Fiske, D.W. e.a. (1970) Planning Research on effectiveness of psychotherapy. *Arch. Gen. Psychiatry*. 22,1 pag. 22-32.
- Fiske, D.W. (1975). The use of significant others in assessing the outcome of psychotherapy. In: Waskov, I.E., Parloff, M.B. *Psychotherapy change measures*. NIMH, Washington.
- Fiske, D.W. (1977). Methodological issues in research on the psychotherapist. In: Gurman, A.S., Razin, A.M. *Effective psychotherapy*. Pergamon, Oxford.
- Fiske, D.W. (1983). The meta-analytic revolution in outcome research. *Journal of consulting and clinical psychology*, 51 (1), pag. 65-70.
- Garfield, S.L., Bergin, A.E. (1978). *Handbook of psychotherapy and behavior change*. John Wiley & Sons, New York.
- Garwick, G., Brintnall, J. (1974). *Proceedings of the second G.A.S. conference*. Minneapolis.
- Garwick, G.B., McCree, S., Brintnall, J. (1976) *Bibliography on Goal Attainment Scaling and Associated Methodologies*, PERC, Minneapolis.
- Additions to the Bibliography on Goal Attainment Scaling and Associated Methodologies, May, 1977.
- Additions to the Bibliography on Goal Attainment Scaling and Associated Methodologies, 1978.
- Garwick, G.B., Brintnall, J.E. (1977). *An introduction to goal attainment scaling, catalogue assisted*. T.A.P.E., Minneapolis.
- Garwick, G., Lampman, Sh. (1978). *Content analysis procedures for goal attainment scaling*.
- Giel, R., Dijk, S., Van Weerden-Dijkstra, J.R. (1978) De nieuwe chronische populatie. *Tijdschrift voor Psychiatrie*, p. 601-609.
- Glass, G.V. (1976). *Primary, secondary, and meta-analysis of research*. San Francisco.
- Greenberg, B.G., Mattison, B.F. (1955). The whys and wherefores of program evaluation. *Canadian journal of public health*, 46, pag. 299.
- Greenfield, T.K. (1983). The role of client satisfaction in evaluating university counseling services. *Evaluation and program planning*, 6 (3,4), pag. 315-321.
- Groot, A.D. de (1961) *Methodologie*. Menton, 's-Gravenhage.
- Grossberg, J.M. en Wilson, H.K. (1965). A correlational comparison of the Wolpe-Lang Fear Survey Schedule and Taylor Manifest Anxiety Scale. *Behaviour research and therapy*, 3, pag. 125-128.
- Hageman-Smit, J. (1976). *De client en zijn hulpverlener, een paar apart*. Samsom. Alphen aan den Rijn.
- Haley, J. (1975). Why a mental health clinic should avoid family *Journal of marriage and family counseling*. 1, pag. 3-13.
- Ham, F. van (1984). Programma evaluatie: iets nieuws of gewoon goed management? In: Schippers, e.a. *Programma-evaluatie in de geestelijke gezondheidszorg*. Swets en Zeitlinger, Lisse.

- Hargreaves, W.A., Clifford, C. (1974). *Current applications of evaluation*, (rapport).
- Hargreaves, W.A. (e.a.) (1977). Short vs long hospitalization: a prospective controlled study. *Archives of gen. psychiatry*, 34, pag. 305-311.
- Harman, H.H. (1967). *Modern factor analysis*. Un. of Chicago press, Chicago.
- Hart, O. van der & Lange, A. (1972) Goal Attainment Scaling, een methode tot evaluatie van psychotherapieën. In: *Maandblad Geestelijke Volksgezondheid*, 27,11 pag. 502-510.
- Haveman, M.J. (1980). Nieuwe langblijvers in psychiatrische ziekenhuizen. *Maandblad voor Geestelijke Volksgezondheid*, p. 642-654.
- Heijs, H. (1984). Programma-evaluatie aan den lijve. In: Schippers, e.a. *Programma-evaluatie in de geestelijke gezondheidszorg*. Swets en Zeitlinger, Lisse.
- Hermans, H.J.M., Tak-van de Ven, J.C.M. (1973). Bestaat er een oorspronkelijke dimensie positief innerlijk welbevinden? *Nederlands tijdschrift voor de psychologie*, 27, pag. 731-754.
- Hersen, M. en Bellack, A.S. (1976). *Behavioral Assessment: a practical handbook*. Pergamon Press, Oxford.
- Hesselink, A.J. (1981). *Metten als geweten*. PUK-symposium 12 juni 1981, Utrecht.
- Hesselink, A.J. & Bolten, M.P. (1986). Kortdurende klinische psychotherapie en behandel-effecten op lange termijn. *Tijdschrift voor Psychotherapie*. 12,1, p.43-53.
- Hodiamont, P. (1986) *Het Zoeken van Zieke Zielen*. Dissertatie Kath. Universiteit. Inst. voor Sociale Geneeskunde, Nijmegen.
- Hoevernaars, J. & van Son, M. (1984). Drie Nederlandstalige zelfrapportage-lijsten in de diagnostiek van depressie. *Gedragstherapie*, 17 (3), pag. 225-239.
- Hofstee, W.K.B. (1980). *De empirische discussie*. Boom, Mepel.
- Honert, J. van den (1841). *Geschiedenis en beginselen der Nederlandsche wetgeving betreffende de gestichten voor krankzinnigen, en de wijze hunner opnemings in en ontslag uit dezelve*. C.G. Sulpke en Gebroeders Diederichs, Amsterdam.
- House, R. (1980). *Evaluating with validity*. Sage Publications, London.
- Hout, M. van den (1984). De wijsneus als niksnut: argumenten voor eenvoud en spaarzaamheid bij programma-evaluatie. In: Schippers, e.a. *Programma-evaluatie in de geestelijke gezondheidszorg*. Swets en Zeitlinger, Lisse.
- Hout, A.C. van den (1985). *Ontslagen psychiatrische patiënten*. Dissertatie, I.T.S., Nijmegen.
- Howard, K.M., Loepp, Ch.E. (1978). *A closer look at client satisfaction as a measure of community mental health program success*. Denver.

- Jaarverslag Geneeskunde Inspectie voor de Geestelijke Volksgezondheid over de periode 1967-1984. Staatstoezicht op de volksgezondheid, Leidschendam.*
- Kazdin, A.E. (1977). Assessing the clinical or applied importance of behavior change through social validation. *Behavior modification*, 1 (4), pag. 427-451.
- Kendell, R.E. (1975). *The rol of diagnosis in psychiatry*. Blackwell Scientific publications, Oxford.
- Kerlinger, F.N. (1967). *Foundations of behavior research*. Holt, Rinehart and Winston Inc. New York.
- Kiesler, D.J. (1977). *Use of individualized measures in psychotherapy and mental health program evaluation research: a review of target complaints, problem oriented record and goal attainment scaling*. report.
- Kiresuk, Th.J. (e.a.).(z.j.) *Service delivery and evaluation from the consumer point of view*. Minneapolis.
- Kiresuk, Th.J., Sherman, R.E. (1968). G.A.S.: a general method for evaluating comprehensive community mental health programs. *Community mental health journal*, 4 (6), pag. 443-453.
- Kiresuk, T.J. (1973). *Goal attainment scaling at an county mental health service, evaluation, vol. 1, 2*. Sage Publications, London.
- Kiresuk, Th.J. (1973). G.A.S. at a county mental health service. *Evaluation*, 1 (2).
- Kiresuk, Th.J., Lund, S.H. (1977). Program evaluation and the management of organizations. In: Anderson, W.F. (e.a.) *Managing human services*. Int. City Management Ass., Washington D.C.
- Kiresuk, T., Sherman, R. (1977). Reply to the critique of GAS. *Social work research and abstracts*, pag. 9-12.
- Kiresuk, T., Sherman, R. (1977). Reply. *Social work research and abstracts*, pag. 47-48.
- Kiresuk, T.J., Lund, S.H. (1978). Goal attainment scaling. In: Attkisson, C.C., Hargreaves, W.A. (ed.) *Evaluation of human service programs*. Academic press, New York.
- Koopmans, G., Bouma, P. (1984). Gebruik van een management informatie systeem bij management en beleidsbeslissingen. In: Schippers, e.a. *Programma-evaluatie in de geestelijke gezondheidszorg*. Swets en Zeitlinger, Lisse.
- Koster, A.M., Wagenborg, J.E.A. (1985) Het follow-up project psychotherapeutische gemeenschappen (3): Enkelvoudige maten voor verandering. *Tijdschrift voor Psychiatrie*. 27,4 pag.272-289
- Koster, A.M. (1986). Het follow-up project psychotherapeutische gemeenschappen (7): cliënten als beoordelaars van de effectiviteit van hun behandeling. *Tijdschrift voor psychiatrie* 28,8 pag. 555-581.
- Kraaiaat, F.W., van Dam-Baggen, C.M.J. (1976). Ontwikkeling van een zelf-beoordelingslijst voor obsessief-compulsief gedrag. *Nederlands tijdschrift voor de psychologie*, 31 (4), pag. 201-211.

- Kraus, G. (1933). *Krankzinnigheid in Nederland*. Proefschrift Groningen.
- Lange, M.M. (1984) *Drie vormen van behandeling in de geestelijke gezondheidszorg*. Swets & Zeitlinger.
- Larsen, J.K. & Werner P.D. (1981). Measuring utilisation of mental health program consultation. In: Ciarlo, J.A. (ed). *Utilisation Evaluation*. Sage Publication, Beverly Hills.
- Larsen, N. *A victory, scoring categories*. P.E.R.C., Minneapolis.
- Lebow, J.L. (1983a). Research assessing consumer satisfaction with mental health treatment: a review of findings. *Evaluation and program planning*, 6 (3,4), pag. 211-236.
- Lebow, J.L. (1983b). Similarities and differences between mental health and health care evaluation studies assessing consumer satisfaction. *Evaluation and program planning*, 6 (3,4), pag. 237-246.
- Lehman, A.F., Zastowny, T.R. (1983). Patient satisfaction with mental health services: a meta-analysis to establish norms. *Evaluation and program planning*, 6, (3,4), pag. 265-274.
- Levine, R.A. e.a. (1981) *Evaluation Research and Practice*. Sage publication, Beverly Hills.
- Lewin, K. (1967). *Resolving social conflicts*. Harper and Row, New York.
- Linden, W.J. van der (1980) *Psychometric contributions to the analysis of criterion-referenced measurements*. Dissertatie Universiteit van Amsterdam.
- Linder-Pelz, S., Epstein, L., Tamir, A. (1983). The meaning of patient satisfaction with prepaid primary health care in Israel. *Evaluation and program planning*, 6 (3,4), pag. 385-394.
- London, P., Klerman, G.L. (1982). Evaluating psychotherapy. *American journal of psychiatry*, 139 (6), pag. 709-717.
- Lord, R.M. & Novick, M.R. (1968). *Statistical Theories of mental test scores*. Addison-Wesley Publ. Comp. Massachusetts.
- Luborsky, L, Chandler, M., Auerbach, A.H. (e.a.) (1971). Factors influencing the outcome of psychotherapy. *Psychological bulletin*, 75, pag. 145-185.
- Luborsky, L., Mintz, J., Auerbach, A. (e.a.) (1980). Predicting the outcome of psychotherapy. Finding of the Penn psychotherapy project. *Archives general psychiatry*, 37, pag. 471-481.
- Luborsky, L., Singer, S., Luborsky, L. (1975). Comparative studies of psychotherapy. *Archives of general psychiatry*, 32, pag. 719-729.
- Luteijn, F., Strarren, J., van Dijk, H. (1975). *Nederlandse persoonlijkheidsvragenlijst, handleiding*. Swets & Zeitlinger, Lisse.
- Lyons Morris, L., Henderson, M., Fitz-Gibbon, C.T. (1979). *The program evaluation kit*. Sage Publications, London.
- Malan, D.H., London, D.M. (1973). The outcome problem in psychotherapy research. *Archives of general psychiatry*, 29, pag. 719-730.

- Marks, J. (1978). Behavioral psychotherapy of adult neuroses. In: Garfield, S.L., Bergin, E.A. *Handbook of psychotherapy and behavior change*. John Wiley & Sons, New York.
- Mastboom, J. (1984). Programma-evaluatie: voorwaarden en mogelijkheden. In: Schippers, e.a. *Programma-evaluatie in de geestelijke gezondheidszorg*. Swets en Zeitlinger, Lisse.
- Mayer, S.E. (1976). *Organizational readiness to accept program evaluation questionnaire*. Minneapolis.
- McGaw, B. (1981). Exploratory data analysis. In: Smith, N.L. (ed.) *New techniques for evaluation*. Sage, London.
- Meehl, P.E. (1954). *Clinical versus statistical prediction*. University of Minnesota Press, Minneapolis.
- Meldman, M. (e.a.) (1976). *The problem oriented psychiatric index and treatment plans*. Kimpton, London.
- Melief, W., Hoekstra, M., Langerak, E. (e.a.) (1979). *Doelen stellen en evalueren*. Samsom, Alphen aan den Rijn.
- Mintz, J., Luborsky, L., Christoph, P. (1979). Measuring the outcome of psychotherapy. Finding of the Penn psychotherapy project. *Journal of consulting and clinical psychology*, 47 (2), pag. 319-334.
- Mintz, J. (1981). Measuring outcome in psychodynamic psychotherapy. *Archives of general psychiatry*, 38, pag. 503-506.
- Mintz, J. (1983). Integrating research evidence: a commentary on meta-analysis. *Journal of consulting and clinical psychology*, 51 (1), pag. 71-75.
- Nguyen, T.D., Attkisson, C.C., Stegner, B.L. (1983). Assessment of patient satisfaction: development and refinement of a service evaluation questionnaire. *Evaluation and program planning*, 6 (3,4), pag. 299-314.
- Norusis, M.J. (1986). *SPSS-X, Advanced statistics guide*. McGraw Hill, New York.
- Orlinsky, D.E., Howard, K.I. (1978). The relation of process to outcome in psychotherapy. In: Garfield, S.L., Bergin, E.A. *Handbook of psychotherapy and behavior changes*. John Wiley & Sons, New York.
- Overwaal (1974 -1985). *Jaarverslagen*.
- Pascoe, G.C. (1983). Patient satisfaction in primary health care: a literature review and analysis. *Evaluation and program planning*, 6 (3,4), pag. 185-210.
- Pascoe, G.C., Attkisson, C.C. (1983). The evaluation ranking scale: a new methodology for assessing satisfaction. *Evaluation and program planning*, 6 (3,4), pag. 335-348.
- Pascoe, G.C., Attkisson, C.C., Roberts, R.E. (1983). Comparison of indirect and direct approaches to measuring patient satisfaction. *Evaluation and program planning*, 6 (3,4), pag. 359-372.
- Pascoe, G.C., Attkisson, C.C., Roberts, R.E. (1983). Relationship of service satisfaction to life satisfaction and perceived well-being. *Evaluation and program planning*, 6 (3,4), pag. 373-384.

- Patton, M.Q. (1978). *Utilization-focused evaluation*. Sage Publications, London.
- Patton, M.Q. (1980). *Qualitative evaluation methods*. Sage Publications, London.
- Patton, M.Q. (1982). *Creative evaluation* Sage publications. London.
- Paul, G.L. (1967). Strategy of outcome research in psychotherapy. *J. of Consulting psychology* 31, p. 104-118.
- Peeters, H.F.M. (1982). Historische fasen in aard en behandeling van geestesziekten. In: Binnenveld, J.M.W., Brinkgreve, C., Lameijn, A.J. *Een psychiatrisch verleden. Uit de geschiedenis van de psychiatrie*. Ambo, Baarn.
- Perloff, R., Perloff, E. & Sussna, E. (1976). Program evaluation. In: Rosenzweig, M.R. & Porter, L.W. (ed.) *Annual Review of psychology*, vol. 27. Palo Alto, Calif.
- Petit, E., Barrett, T. *Using goal attainment scaling in managing a community mental health center*. Bethesda M.H.C., Denver.
- Prazak, J. (1978a). *Carf program evaluation project participant workbook*. Second session. Walker and associates inc., Minneapolis.
- Prazak, J. (1978b). *A guide to using your evaluation system and making it work for you*. Walker and associates inc., Minneapolis.
- Prazak, J. (1978). *Carf program evaluation project participant workbook*. Walker and associates evaluation systems, Minneapolis.
- Roberts, R.E., Attkisson, C.C. (1983). Assessing client satisfaction among hispanics. *Evaluation and program planning*, 6 (3,4), pag. 401-414.
- Romijn, C. (1984). Evaluatie van gezinstherapie in de ambulante verslavingszorg. In: Schipper, G. *Programma-evaluatie in de geestelijke gezondheidszorg*. Swets & Zeitlinger, Lisse.
- Romme, M.A.J. (1984). *Geestelijke gezondheidszorg. Voorzieningen, werkwijzen, doelgroepen, medewerkers, ontwikkelingen, financiering, omvang*. Stafleu, Alphen aan de Rijn.
- Rossi, P.H., Wright, S.R., Freeman, H.E. (1979). *Evaluation (a systematic approach)*. Sage Publications, London.
- Ruiter, D. (1979). *Management van non-profit organisaties*. Samsom, Alphen aan den Rijn
- Rutman, L. (1977) *Evaluation Research Methods: a basic Guide*. Sage publications, London.
- Rutman, L.S., Wholey, J.S. (1980). *Planning useful evaluations (evaluability assessment)*. Sage Publications, London.
- Schagen, S. (1983). *Het effect van psychotherapie Meetbaarheid en Resultaten*. Van Loghum Slaterus, Deventer.
- Schagen, S. (1985). *De praktijk van de psychotherapie*. Van Loghum Slaterus, Deventer.
- Schmidt, H. (1974) Goal Attainment Scaling III, Een methode ter evaluatie van psychotherapieen? In: *Maandblad Geestelijke Volksgezondheid*, 29,3 pag. 139-145.

- Schnabel, P. (1984a). Onderzoek en ordening in de geestelijke gezondheidszorg op regionaal niveau. In: Etten, G. van. *Beleidsinformatie en onderzoek in de gezondheidszorg. Kwartaalschrift leiding en organisatie in de gezondheidszorg*, 1984/1, pag. 100-130.
- Schnabel, P. (1984b). Programma-evaluatie: een nadere kennis-making bij het scheiden van de markt. In: Schippers, e.a. *Programma-evaluatie in de geestelijke gezondheidszorg*. Swets en Zeitlinger, Lisse.
- Schutz, H. (1981). Gedragstherapie bij IMP-clienten. *De psycholoog*, 16 (10), pag. 690-701.
- Schuurmans Stekhoven, J.H. (1922). *Ontwikkeling van het krankzinnigenwezen in Nederland 1813-1914*. Algemeene Landsdrukkerij, 's-Gravenhage.
- Schwab, M.E., Stone, K. (1983). Conceptual and methodologic issues in the evaluation of children's satisfaction with their mental health care. *Evaluation and program planning*, 6 (3,4), pag. 283-290.
- Seaberg, J.R., Gillespie, D.F. (1977). GAS: a critique. *Social work research and abstracts*, pag. 4-9.
- Seaberg, J.R., Gillespie, D.F. (1977). Reply. *Social work research and abstracts*, pag. 46-47.
- Seaberg, J.R., Bill, M.A. (1978). Evaluation of broad-scale service networks. *Social work research abstracts*, pag. 19-24.
- Seidel, H. (1982) *Aristoteles' meta physik*. Erster Halbband: Bucher I(A)-VI(E) Felix Meiner Verlag, Hamburg.
- Shapiro, D.A. & Shapiro, D. (1982). Meta-analysis of comparative Therapy Outcome Studies: A replication and Refinement. *Psychological Bulletin*. 92,3 pag. 581-604.
- Shapiro, D.A., Shapiro, D. (1983). Comparative therapy outcome research: Methodological implication of meta-analysis. *Journal of consulting and clinical psychology*, 51 (1), pag. 42-53.
- Sherman, R. (1974). The content validity argument for the validity of the goal attainment score. *Newsletter P.E.R.C.*, november/december.
- Sherman, R. (1977). Will GAS solve the problems of program evaluation in the mental health field? In: Coursey, R.D. (ed.) *Program evaluation for mental health methods, strategies and participants*. Grune and Stratton.
- Sijben, N. (1977a) *P.O.P.E. memo 1: 1e interim nota 1976/1977*. Kath. Universiteit, afd. Psychiatrie, Nijmegen.
- Sijben, N. (1977b) *P.O.P.E. memo 2: Overzicht statistische gegevens patientenpopulatie 1974, 1975, 1976*. Kath. Universiteit, afd. Psychiatrie, Nijmegen.
- Sijben, N. (1977c) *P.O.P.E. memo 3: Follow-up onderzoek 1974-1977*. Kath. Universiteit, afd. Psychiatrie, Nijmegen.
- Sijben, N. (1979a) *P.O.P.E. memo 4: Program-Evaluatie in U.S.A.* Kath. Universiteit, afd. Psychiatrie, Nijmegen.
- Sijben, N. (1979b) *P.O.P.E. memo 5: Follow-up onderzoek 1977-1979*. Kath. Universiteit, afd. Psychiatrie, Nijmegen.

- Sijben, N. (1979c) *P.O.P.E. memo 6: Project Overwaal Programma Evaluatie - subsidieaanvraag*. Kath. Universiteit, afd. Psychiatrie, Nijmegen.
- Sijben, N. (1979d) Reactie op "Hulpverlening aan astmatische kinderen en hun gezinnen" van Veugelers e.a. In: *Maandblad Geestelijke Volksgezondheid*, 7/8.
- Sijben, N., Zwaan, E.J., Zwart, F.M. (1979e). Gedragstherapeutische middelen. In: *Handboek gedragstherapie*, deel D, pag. D3-2 - D3-8.
- Sijben, N. (1980a). Program evaluatie. *Bulletin vereniging voor gedragstherapie*, 13 (1), pag. 5-20.
- Sijben, N. (1980b) *Het meten van therapie resultaten*. Lezing Najaarsconferentie Vereniging voor Gedragstherapie, Woudschoten.
- Sijben, N. (1981). *P.O.P.E. Interim rapport nr. 1, Memo 7*. Kath. Universiteit, afd. psychiatrie, Nijmegen.
- Sijben, N. (1982). *P.O.P.E. Interim rapport nr. 2, Memo 9*. Kath. Universiteit, afd. psychiatrie, Nijmegen.
- Sijben, N. (1983). *P.O.P.E. Interim rapport nr. 3, Memo 10*. Kath. Universiteit, afd. psychiatrie, Nijmegen.
- Sijben, N. (1984a). Programma-evaluatie, wat is dat? In: Schippers, e.a. *Programma-evaluatie in de geestelijke gezondheidszorg*. Swets en Zeitlinger, Lisse.
- Sijben, N. (1984b). Programma-evaluatie gepeild. In: Schippers, e.a. *Programma-evaluatie in de geestelijke gezondheidszorg*. Swets en Zeitlinger, Lisse.
- Sijben, N. (1984c). *P.O.P.E. Rapportage 1980-1984. Memo 11*. Kath. Universiteit, afd. Psychiatrie, Nijmegen.
- Sloane, R.B. e.a. (1975) *Psychotherapy versus behavior therapy*. Cambridge Harvard University Press.
- Smith, M.L., Glass, G.V. (1977). Meta-analysis of psychotherapy, outcome studies. *American psychologist*, pag. 752-760.
- Smith, M.L. & Glass, G.V. (1977) Meta-analysis of Psychotherapy Outcome Studies. *American Psychologist*.p. 752-761.
- Smith, N.L. (1981) *New Techniques for Evaluation*. Sage publications London.
- Soelling, M.E., Newell, T.G. (1983). Effects of anonymity and experimenter demand on client satisfaction with mental health services. *Evaluation and program planning*, 6 (3,4), pag. 329-334.
- Soudijn, K.A. (1976). *Dilemma's in sociaal wetenschappelijk onderzoek*. Boom, Meppel.
- Soudijn, K.A. (1982). *Kwaliteit van psychotherapie*. Boom, Meppel.
- Soudijn, K.A. (1984). *Conferentie - Psychotherapie: de dialoog tussen onderzoek en praktijk*. Internationale school voor wijsbegeerte, Leusden.
- Spitzer, R.L., Williams, J.B.W., Skodol, A.E. (1980) DSM III: The major achievement and an overview *The American Journal of psychiatry*, 137,2 p. 151-164.
- Stam, F.C. (1975). The Netherlands. In: Howells, J.G. & Psych, F.R.C. *World history of psychiatry*. Baillière Tindall, London.

- Stokking, K.M. (1984) *Interpreteren en evalueren*. Van Loghum Slaterus, Deventer.
- Strupp, H.H., Bergin, A.E. (1969). Some empirical and conceptual bases for coordinated research in psychotherapy. *International journal of psychiatry*, 7 (2), pag. 18-90.
- Strupp, H.H. (1976). Themes in psychotherapy research. In: Claghorn, J.L. (red.), *Successful therapy*. Brunner/Mazel, New-York. general psychiatry, 37, pag. 947-954.
- Strupp, H.H. (1981). Clinical research, practice and the crisis of confidence. *Journal of consulting and clinical psychology*, 49, pag. 216-219.
- Suchman, E.A. (1967). *Evaluative Research: Principles and practice in public service and social action programs*. Russel Sage foundation, New York.
- Swanborn, P.G. (1971). *Aspecten van sociologisch onderzoek*. Boom, Meppel.
- Swanson, R.M. (e.a.) (1978). *Dimensions of mental health treatment goals: a tri informant conception*. Denver.
- Taylor, J.A. (1953). A personality scale of manifest anxiety. *Journal of abnormal and social psychology*, 48 (2), pag. 285-290.
- Thompson, M. (1980). *Benefit-cost analysis for program evaluation*. Sage Publications, London.
- Uberla, K. (1971). *Faktorenanalyse*. Springer Verlag, Berlin.
- Van Dale (1984) *Groot Woordenboek der Nederlandse Taal*. Van Dale Lexicografie, Utrecht.
- Vandermeersch, P. (1984). *Doktoren aan de ziel. Historische aspecten van de verhouding tussen religie en psychiatrie*. Symposiumverslag, NcGv, Utrecht.
- Veen, J. van de, Brinkman, W. (1983). De moeilijkheid van de client en de resultaten van de psychotherapie. *Tijdschrift voor psychotherapie*, 9 (3), pag. 116-129.
- Veer, J. van der (1974) Goal Attainment Scaling IV, Doelenkontraakt als evaluatiemethode. *Maandblad Geestelijke Volksgezondheid*, 29,10 pag. 469-480.
- Veerman, J.W. (1983) Psychotherapie-onderzoek in beweging: van demonstratie naar registratie. *Tijdschrift voor psychotherapie*. pag. 230-236.
- Verslag over den staat der gestichten voor krankzinnigen over de perioden 1844-1846 (nr. 1) tot en met 1882-1884 (nr. 13).*
- Verslag over het staatstoezicht op krankzinnigen en krankzinnigengestichten en over den staat dier gestichten in de perioden 1885-1887 (nr. 14) tot en met 1932-1936 (nr. 28).*
- Veugelers, P.P. e.a. (1978) Hulpverlening aan astmatische kinderen en hun gezinnen. *Maandblad Geestelijke Volksgezondheid*, 33,7/8, pag. 524-536.
- Vijselaar, J. (1983). Een eerste inspectie, over de inspectie in de negentiende eeuw. *Maandblad geestelijke volksgezondheid*, 38 (6/7), pag. 650-665.
- Visser, A. Ph. (1978). Het innerlijk welbevinden van ziekenhuispatienten. *Gedrag*, 6, pag. 402-418.

- Walker, R.A. (1972). The ninth panacea: program evaluation. *Evaluation*, pag. 45-53.
- Ware Jr., J.E., Snyder, M.K., Russell Wright, W. e.a. (1983). Defining and measuring patient satisfaction with medical care. *Evaluation and program planning*, 6 (3,4), pag. 247-264.
- Ware Jr., J.E., Davies, A.R. (1983). Behavioral consequences of consumer dissatisfaction with medical care. *Evaluation and program planning*, 6 (3,4), pag. 291-298.
- Warries, E. (1970) Het relatief meten van leerprestaties in het onderwijs. *Ned. Tijdschrift voor de Psychologie*, 25, pag. 429-439.
- Waskov, I.R., Parloff, M.B. (red.) (1975). *Psychotherapy change measures*. NIMH, Washington.
- Weerden-Dijkstra, J.R. van (1976) Een aantal karakteristieken van het chronische bestand. *Tijdschrift voor Psychiatrie*, p. 120-132.
- Wilde, G.J.S. (1963). *Neurotische labiliteit gemeten volgens de vragenlijstmethode*. Van Rossen, Amsterdam.
- Willems, L.F.M., Tuender-de Haan, H.A., Defares, P.B. (1973). Een schaal om sociale angst te meten. *Nederlands tijdschrift voor de psychologie*, 28, pag. 415-422.
- Willer, B., Howard, D. (1974). Development of a computer assisted progress record for psychiatry. *Goal attainment review*, 1, pag. 35-43.
- Wilson, N.C., Ellis, R.H. (z.j.) *Atgon: goals, interview manual, team manual sample forms and methods definitions*. Fort Logan mental health center.
- Wilson, N.C. (1973a). The automated tri-informant goal-oriented progress note. *Journal of community psychology*, 1 (3), pag. 302-306.
- Wilson, N.C. (e.a.) (1973b). *Treatment effectiveness, program evaluation and Atkon*. Fort Logang mental health center, august.
- Wilson, N.C. (1974). *The Atgon approach to program evaluation*. Denver, paper.
- Wilson, N.C., Ellis, R.H. (1975). *Atgon: a multipurpose information system for adult psychiatry*. Report 1-5, march.
- Wilson, N.C. (1977). *Treatment outcome: a feasibility pilot study*. Division of mental health, Denver.
- Winer, B.J. (1962). *Statistical principles in experimental design*. McGraw-Hill, London.
- Wit, C.A.M. de (1975) Goal Attainment Scaling V. *Maandblad Geestelijke Volksgezondheid*, 30,4 p. 219-225.
- Wojciechowski, F.L. (1984). *Double blind research in psychotherapy*. Swets & Zeitlinger, Lisse.
- Wolpe, J. & Lang, P.J. (1964). A Fear Survey Schedule for use in behaviour therapy. *Behavior research and therapy*, 2, pag. 27-30.
- World Health Organisation (1978). *Manual of the international classification of diseases injuries and causes of death*, 9th revision, (1). Geneva.

- Worthman, P.M. (1981). *Methods for evaluating health services*. Sage Publications, London.
- Zung; W.W.K. (1965). A self rating depression scale. *Archives of general psychiatry*, 12, pag. 63-70.
- Zwaan, A. de (1983). *Met de prop in de mond - over de vrijheid van onderzoek*. Vrij Nederland, 1-10-1983.

Curriculum Vitae

No Sijben werd geboren op 24 oktober 1946 te Hoensbroek (L.). Na het Gymnasium β studeerde hij psychologie te Utrecht. In mei 1973 behaalde hij het doctoraal examen in twee hoofdrichtingen: klinische (differentiële) psychologie en methodologie/statistiek, met als bijvak psychopathologie.

Van 1967 tot 1974 was hij student-assistent op het Instituut voor Clinische en Industriële Psychologie, afd. testmethodiek en testarchief. Gedurende twee jaar was hij ook student-assistent op de afdeling Methoden en Technieken.

Na zijn afstuderen trad hij in dienst bij het Centraal Instituut voor Toetsontwikkeling (C.I.T.O., Arnhem), waar hij als methodoloog werkzaam was.

Sinds 1976 is hij research-coördinator van de afdeling psychiatrie van de K.U. Nijmegen. Naast beleidsmatige werkzaamheden begeleidt hij onderzoeksprojecten van medewerkers en (keuze-)stagiaires en verzorgt hij onderwijs o.a. in methodologie en psychodiagnostiek.

In deeltijd was hij van 1976-1984 gedetacheerd op Kliniek Overwaal (Lent) voor het opzetten van een systeem voor de (permanente) evaluatie van de behandeling. Over deze periode handelt deze dissertatie.

Stellingen behorende bij de dissertatie:

Omzien naar weldoen

Programma-evaluatie in theorie en praktijk

van N. Sijben.

- 1 In tegenstelling tot de gebruikelijke wetgeving was de zgn Krankzinnigenwet van 1841 zijn tijd ver vooruit, met een doelstelling waaraan vele jaren lang de praktijk in positieve zin aangepast diende te worden. De eerste 'inspecteurs der gestichten voor krankzinnigen' waren de eerste programma-evaluatoren in de GGZ: zij toetsten permanent de praktijk aan deze doelstelling, en met succes, hun feedback had resultaat.

Dit proefschrift, hoofdstuk 1

- 2a Het principe '*No cure, no pay*' moet voor de (G)GZ niet bij voorbaat worden verworpen. Met programma-evaluatie kan deze 'cure' worden aangetoond.

Dit proefschrift, hoofdstuk 3 & 11

- 2b De huidige stringente bezuinigingen in de (G)GZ leiden overigens tot een beleid waarbij het principe '*No pay, no cure*' in ieder geval meer en meer opgaat.

o a Concept Jaarplan 1987, St. Radboudziekenhuis, Nijmegen

- 3 In tegenstelling tot de 'profit'-sector dient de GGZ als 'non-profit'-instelling dusdanig kwalitatief werk te leveren dat haar cliënten niet (te) lang blijven en niet meer terugkomen (uit positieve overwegingen). De toename van het aantal langdurig verblijvende (chronische) patiënten en de stijging van het aantal heropnamen in Algemeen Psychiatrische Ziekenhuizen is hiermee in strijd.

Dit proefschrift, hoofdstuk 1 & 11

- 4 De 'moeilijkheidsgraad van een cliënt' is geen absoluut maar een relatief begrip. Onderzoek naar zgn. moeilijk behandelbare patiënten/cliënten dient zich dan ook vooral te richten op de behandelaar en de gebruikte behandelingsmethoden, in plaats van op patiëntenkenmerken (cf. Van de Veer & Brinkman). Dit onderzoek heeft tevens het risico dat als 'moeilijk' geclassificeerde patiënten (nog) minder kans maken behandeld te worden.

Van de Veer en Brinkman, (1983) De moeilijkheid van de cliënt en de resultaten van psychotherapie. *Tijdschrift voor psychotherapie* 9 3, p. 116-129.

Dit proefschrift, hoofdstuk 9

- 5 In psychiatrisch epidemiologisch onderzoek wordt o a. gezocht naar risicofactoren en -indicatoren voor casuïstiek, door verbanden te leggen tussen het feit of iemand al dan niet een psychiatrische casus is en eenvoudig traceerbare demografische kenmerken als geslacht, leeftijd, burgerlijke staat, etc. Deze relaties hebben empirisch en theoretisch weinig betekenis.

Hodiamont, P. P. G., Peer, P. G. M., & Sijben, A. E. S. (1986) Demografische aspecten van psychiatrische problematiek in de gezondheidsregio Nijmegen. In: Hodiamont, P. P. G. *Het zoeken van zieke zielen*. Proefschrift, Nijmegen.

- 6 In de Nederlandse vertaling van de Mini Mental State, ter bepaling van demantie, is het woord 'world' dat de onderzochte van achter naar voren moet spellen, vertaald door 'wereld'. Hieruit blijkt dat de vertaler/ster de testopgave niet begrepen heeft. Onderzoek naar de diagnostiek van demantie verdient grotere zorgvuldigheid.

Een spellingsopdracht voor het woord 'worst' is hier beter op zijn plaats.

7. Ten aanzien van biologische parameters in de psychiatrie weten we nu dat we minder weten dan we dachten te weten. Ook dat is vooruitgang in de wetenschap.

cf. Hoes, M.J.A.J.M. (1986). Biological markers in psychiatry.
Acta psychiatrica belgica 86,3 p. 220-241.

8. Het bij de faculteit der geneeskunde (K.U. Nijmegen) gebruikte systeem voor correctie voor raden bij meerkeuze tentamens, waarbij punten worden afgetrokken voor onjuist beantwoorde vragen:
- a. heeft geen psychometrische voordelen vergeleken met ongecorrigeerde tentamenscores;
 - b. straft ten onrechte foute antwoorden op basis van onjuiste informatie;
 - c. appelleert te veel aan het risico-nemend vermogen van de student;
 - d. meet naast kennis ook persoonlijkheidsfactoren van de student.
- Bij de beantwoording van deze vierkeuzevraag hoeft niet geraden te worden, omdat alle antwoorden juist zijn. Een dergelijke correctieprocedure is derhalve ongewenst.

Borgesius, T.G. (1978). *Een empirisch onderzoek naar het correctie-voor-raden-scoringssysteem*. I.O.W.O. K.U. Nijmegen.

Diamond, J. & Evands, W. (1973). The correction for guessing.
Review of education research. 43,2 p. 181-191.

9. De reeds in 1637 door Descartes beschreven mogelijkheden (inclusief de spraak) en de grenzen van een humanoïde robot, zijn ook nu nog geldig.

Descartes, R. *Over de methode*. (Ned. vertaling 1977) Boom, Meppel.

10. Uit onderzoek is gebleken dat een volgens het Jenaplan ingerichte basisschool een goede of soms zelfs betere voorbereiding is voor het traditionele voortgezette onderwijs dan een 'gewone' basisschool.

Sijben, N. Het Jenaplan onder de LOEP.
Mensen kinderen nov. 1986, nr. 2.2.

11. De bedenker van het eerste evaluatie-onderzoek, Daniël (ca. 600 v. Chr.), dient te worden uitgeroepen tot de patroon van programma-evaluatoren. Ook zijn verdere levensloop kan hen tot voorbeeld strekken: hem werd het vuur na aan de schenen gelegd en hij werd voor de leeuwen geworpen. Beide situaties wist hij echter ongeschonden te doorstaan.
Fatalistische evaluatoren zullen evenwel Socrates prefereren.

12. Het primitieve idee van een cyclisch en eeuwigdurend tijdsverloop wordt goed weergegeven door de vraag van de driejarige Paul Zijben: 'Mama als alle dagen over zijn, komt dan weer vroeger?'.

13. S.D.I. is een slag in de lucht of *wordt* het.
Een slag in de lucht is hoe dan ook *mis*.

11

12

13

14

15

16

17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100

101

102

103

104

105

106

107

108

109

110

111

112

113

114

115

116

117

118